CA2ALHR
72P61

CA2 ALHR 1972P61
Projections of Alberta School Enrollment
s: Methodologies and Their Applica 1
3 3398 00136 5047

PROJECTIONS OF ALBERTA SCHOOL ENROLLMENTS: METHODOLOGIES AND THEIR APPLICATIONS

LIBRARY
VAULT

Digitized by the Internet Archive
in 2026 with funding from
Legislative Assembly of Alberta - Alberta Legislature Library

LEGISLATIVE LIBRARY
AUG 14 1980
EDMONTON

FOR LIMITED DISTRIBUTION ONLY

This report has been submitted to the Human Resources Research Council by Mr. Michael Grace as the final report for the commissioned study:

Projections of Alberta School Enrollments:
Methodologies and Their Applications

It has not been revised or edited and may not be reproduced in any form without the permission of the Council.

PROJECTIONS OF ALBERTA SCHOOL ENROLLMENTS: METHODOLOGIES AND THEIR APPLICATIONS

Michael Grace
John Hanson
Kyung Bay
Dave Flathman
Anne-Marie Decore

This is a commissioned research project of the Alberta Human Resources Research Council, Edmonton 62, Alberta

Digitized by the Internet Archive
in 2026 with funding from
University of Alberta Library

https://archive.org/details/ableg_33398001365047

TABLE OF CONTENTS

		Page
CHAPTER I:	INTRODUCTION	1
CHAPTER II:	LITERATURE REVIEW	10
CHAPTER III:	DATA BASE	20
	3.1 Data Base	20
	3.1.1 Process of the System	21
	3.1.2 Data Format	24
	3.2 Accounting Matrices	26
	3.2.1 Cross-Sectional Tables	27
	3.2.1.1 Description of Age-Grade Tables	28
	3.2.2 Longitudinal Tables	29
	3.2.2.1 Description of Cohort Tables	32
	3.2.3 Relationship of Age-Grade Tables to Cohort Year Tables	34
	3.3 Flow Matrices	35
	3.3.1 Contingency Tables Derived from the Multidimensional Data	39
	3.3.1.1 Year Wise	40
	3.3.1.2 Age Wise	40
	3.4 Models	44
CHAPTER IV:	DATA ANALYSIS	46
	4.1 Introduction	46
	4.2 Polynomial Models	47
	4.2.1 Polynomial Regression on Cohort Year	48

Page

4.2.2 Polynomial Regression on Year 49
4.2.3 Polynomial Regression on Age 49
4.2.4 Polynomial Regression on Age, Grade, and Year 49
4.2.5 Discussion and Comparison 50
4.3 Auto-Regressive Models 65
4.3.1 Two- and Three-Parameter Regression . . 65
4.3.2 Two-Cell Average 74
4.3.3 Auto-Regressive Model 77
4.3.3.1 Auto-Regression Through Cohort Year 78
4.3.3.2 Auto-Regression Along Cohort Diagonals. 78
4.3.4 Giant Regression 79
4.4 Age and Grade Distributions 80

CHAPTER V: PROJECTION BY MULTI-CHANNEL TIME SERIES ANALYSIS . . 86
5.1 Introduction 86
5.2 Model 87
5.3 Assumptions of the Model 91
5.4 Results 94
5.4.1 Order of the Scheme 94
5.4.2 Filters 96
5.4.3 Projections of Enrollment 100

CHAPTER VI: THE MARKOV CHAIN MODEL 106
6.1 Introduction 106
6.2 The Model 107
6.3 The Data and Parameter Estimation 111
6.4 Absorbing Markov Chains 121
6.5 Projections 128

Page

CHAPTER VII: RANDOM WALK SIMULATION 135

7.1 Introduction 135

7.2 Assumptions 136

7.3 Description of the Model 138

7.3.1 Exogenous Variables 139

7.3.2 Status Variables 140

7.3.3 Parameters 142

7.3.4 Endogenous Variables 144

7.3.5 Inputs 145

7.3.6 Outputs 146

7.4 Methodology 148

7.4.1 Transition Rules 151

7.4.2 Identities and Operating Characteristics 154

7.4.3 Summary of Inputs 158

7.5 Projections of Enrollment 163

CHAPTER VIII: COMPARING THE MODELS 176

CHAPTER IX: DISCUSSION AND RECOMMENDATIONS 185

9.1 Data Refinement 185

9.2 Individual vs Aggregate Data 186

9.3 Age Cohort vs Pseudo Age Cohorts 187

9.4 Parameter Refinements 189

9.5 Other Variables 191

9.6 Choosing Among Techniques 192

Page

APPENDIX A: CONVERSION OF THE ALBERTA SCHOOL ENROLLMENT DATA 1945-46 to 1969-70 194

A.1 Introduction 194

A.2 The Conversion Procedure 194

APPENDIX B: SECTION I: PARAMETER ESTIMATION 203

B.1 Parameter Estimation 203

B.1.1 Example of a Portion of a Cohort Table 204

B.1.2 Notation and Definition 205

B.1.3 Method 206

B.1.4 Alternative Methods 211

APPENDIX B: SECTION II: ESTIMATION OF NEW EXTRANTS . . . 212

B.2 Estimation of New Entrants 212

BIBLIOGRAPHY 220

ABSTRACT

The purpose of the study is to examine projection methodology and to analyze data in Alberta's school system. As a secondary product we make enrollment projections to 2005 by several methods and compare the results. These should be used with caution.

A review of the literature is provided along with detailed statements concerning the data. Complete appendixes of results and computer programs are provided separately.

ACKNOWLEDGMENTS

A project of this size usually requires considerable co-operation, and this one was no exception. The Alberta Department of Education and the Division of Educational Research at the University of Alberta assisted in data collecting and provided computer facilities. We especially appreciate the many long hours Jenny Bererton and Jane Percy spent in compiling information and organizing data. And, as always, June Talpash accomplished the immense typing chores in her usual competent and cheerful manner. To them, and to everyone else who made this report possible, our sincere thanks.

CHAPTER I

INTRODUCTION

This study was undertaken to examine projection methodology and to analyze enrollment data in Alberta's grade one to twelve school system for the purpose of making enrollment projections. Actual forecasts made for the next three decades are a function of age, grade, and sex. This is done by using the existing data with consideration of external forces acting on the system such as births, migration, and deaths. These forecasts are made by a number of models giving us different projections under varying conditions. The models used are mathematical ones that have been developed to describe the location of students in the school system. Specifically we:

1. Make projections of grade school enrollment in Alberta for every year up to and including the school year 2004-05.

2. Develop a variety of models as a package to allow users to make projections of their own school system enrollments under varying situations and assumptions.

Since the inception of this study on projecting school enrollments for Alberta the following has been done:

1. A large data collection program was undertaken involving school enrollments by age, grade and sex for the last 25 years in Alberta and the last 16 years in Puerto Rico. The Puerto Rican data has proved useful in checking results obtained from the Alberta

data. All this data was keypunched, verified, sorted, and placed on magnetic tape and refined.

2. Data was collected from various school systems in Canada and the United States pertaining to passing rates, failure rates, dropouts, mortalities, returning dropouts, double promotions, and graduates. This data has assisted us in the selection of similar parameters to be used in conjunction with the Alberta data. Many of these parameters were calculated and refined on the basis of age, grade, and sex.

3. The method of successive subtraction was developed as a means of extracting parameters from the enrollment data.

4. Data was collected from a variety of census sources on births, deaths, migration, and population. Future trends in each were estimated and later used in our major projection models.

5. A comprehensive bibliography of enrollment projection methodology has been collected that is applicable to both schools and universities.

6. Patterns of student movement through the educational system have been studied.

7. A method of making school enrollment data comparable by age in Alberta has been developed to solve the problem of the use of different attendance reporting dates by the Department of Education.

8. A large number of linear and nonlinear models were run on the Alberta data to find any trends that existed in the data.

9. Enrollment projections have been made by time series models, polynomial methods, Markov chain, and simulation. Refinements have been made in the major projection models so that in most cases new parameters and/or variables can be easily incorporated into the models. This allows the user to revise and update his projections as conditions change.

The data base consists of the aggregate school enrollment figures reported by the Alberta Department of Education. These figures are the result of each school district tabulating their enrollment once a year and classifying students by age, grade, and sex. The present "date" of enrollment is September 30 which indicates the school district totals for that given year. In recent years "opportunity classes" have been on the increase which are tabulated separately. We have distributed these classes proportionately across grades in order to account for all students enrolled in the province. The margin of error involved here would be very slight as only a small number of opportunity students are enrolled in the system. As time goes on we would expect the effect of "opportunity classes" to increase, tending to invalidate our assumptions, and in a similar vein, any movement towards non-graded systems would diminish the validity of our grade estimates.

The original enrollment data is supplied in an age-grade-sex table form. Only a few minor changes have to be made to convert the data into the form needed to make projections. The data was keypunched onto coded cards in a systematic format and was then transferred onto magnetic tapes. Provisions have been made for the

addition of new enrollment data. The Alberta Department of Education supplies the data with the district variable already aggregated. Using the raw data the original age-grade tables can be produced along with variations such as cohort summary tables.

Data used in estimating parameters such as pass rates, failure rates, dropouts, births, etc., have been collected from a variety of sources such as the Alberta Department of Education, United States Office of Education, Alberta Bureau of Statistics, Dominion Bureau of Statistics, Saskatchewan Department of Education, and Manitoba Department of Education. Usually the data was not immediately applicable and transformations were often necessary to put the parameters into the form desired. Actual figures were generally put into the form of distributions which are appropriate for stochastic models. For some parameters information was not available and even if it was, it seemed inappropriate for the Alberta school system. As a consequence, a method has been developed that will estimate parameters internal to the educational system.

In any year enrollment data may be collected in terms of numbers of students by age, grade, and sex. Collectively we form a cross-sectional data base that consists of a contingency table by placing the number of individuals, identical with respect to all variables, in the same category. A compilation of this data over a period of years yields longitudinal data in the form of cohort tables. Thus any individual may appear in a chort table many times.

The main purpose in investigating cross-sectional (age-grade tables) and longitudinal (cohort tables) data is to learn

about the processes and factors inherent in movements within the system so that projections can be made. The procedure we follow consists of exploring the system in a variety of ways and then using the results to make projections. The first stage consists of fitting various models to both age-grade and cohort tables. This has a two-fold purpose: (1) to discover those models which are most useful in cleaning or smoothing the data, and (2) to determine which models, if any, exhibit useful trends. The second stage is the collection of data that will be useful in forming parameter distributions and to develop methods of deriving suitable parameters where none are available. The third stage is the development of models, regression and time-series, that do not require parameters for projection purposes. The fourth stage is to develop major models, Markov and random walk simulation, that require internal and external parameters and can be updated, and the fifth stage is comparing and contrasting the models and discussing the results obtained.

The terms projection, prediction, and forecasting are often used synonymously. We restrict our use of the term prediction to making estimates from a process over some time interval for which a data base exists. This often is the method employed to make estimates of parameters. Projections are estimates of the future data base one or more steps along the time interval. Forecasting is a less precise term and is more subjective than projections. Forecasts are unconditional statements of opinion not necessarily based on theoretical laws.

Projections by extrapolation are extensions into the future of time series established in the past. Extrapolations are more reliable in regard to the trend of the series than in regard to the fluctuations around the trend.

The choice of models is determined by the nature of the system, what we want to know, the supply of basic information, and the level of detail at which we wish to work. Our attention is focused on models that are to be used for cleaning the data, determining trends and tendencies, and making projections. Cleaning (or smoothing) is a concept that involves the removal or reduction of discrepancies in the data base. The major cleaning problem involved in the Alberta data was to smooth out the jumps in the data base brought about by the different attendance reporting dates used in Alberta over the past twenty-five years. The smoothed data allows us to apply simpler and more realistic models in determining trends in the variables.

The principles of multiple linear regression underlie the various regression, polynomial, and auto-regression models used in examining the data for trends. Some of the models explored in data analysis can also be used with varying degrees of success in making projections. We develop and then make projections using a giant polynomial model, and the roots of our time-series projections lie in the auto-regressive models. The regression models are grouped together and compared. It is from these analyses that we ascertain the gradual shifts of students progressing through the grades over the years. The strong relationship between grade and age is also

apparent as is the tendency for students to stay in school longer. The main benefits derived from data analysis are the "feeling" developed from working with a large body of data and the ideas generated that aided us in building our major projection models.

Time series are noted as being successive measurements of the value of a variable over a period of time. Any time series may exhibit trends that are any or all of seasonal, cyclical, extraordinary or general. We consider the investigation of all but the last type of trend to be beyond the scope of this work and consequently our major time series effort was on the general trend. This effort was a quantitative investigation of the structure of an educational system based on a discrete rather than a continuous analysis. The discrete form seems to be called for here as most systems work on lock step intervals (graded system) that take place at specified points in time although many of the external inputs into our system such as births and mortalities are of a continuous nature. Even such outputs as graduations may follow a logistic growth curve and care must be taken to preserve such features. Such factors make us believe that the overall approach should be a dynamic one.

Parameters are developed from within the confines of the data base and are also obtained from outside sources and adapted to the Alberta school system. These parameters are classified and discussed as being external or internal to the system in question. External parameters such as births, mortalities, and net migrations are exogenous and as such lie on the outside of the

system relationship which we are specifically studying. Internal parameters such as dropouts and failures are contained with the system and are considered to be endogenous. The accuracy of the projections we obtain are conditional upon the validity of our parameter estimates and models adopted. For example, it is obvious that any changes in the birth rate would eventually affect the estimated school enrollment. Equally obvious would be the endogenous effect of changes in school administration policy requiring attendance of all students until they reach the age of 18.

The major projection models used in later chapter, Markov chain and random walk simulation are systematically built from theoretical structures. Where possible we test them for internal consistency and subject certain aspects of them to empirical tests. We realize that the concepts used tend to be abstract and hypothetical in nature but we endeavour to formulate them so as to exhibit a close relation to reality. The models that we finally settle on are the products of a great deal of objective and subjective study on numerous multirelational models.

Markov chains exist if the probabilities for passing from one state to the next are determined by the existing state of the system. We represent the flow of students through the system as a Markov process. In this manner we envision the educational process as a series of events subject to random fluctuations over a period of time. The development of the system depends only on its state at the present time and inputs to the system, in the form of new entrants, at given points of time in the future. The Markov model

allows us to investigate the structure of the educational system and to predict future enrollment distributions of students.

The random walk model is the basis for the simulation of the educational process with appropriate experiments on the model to determine its value. The simulation permits the random generation of data which may be used as a basis for formulating hypotheses about the system under observation. Experimentation with a number of alternative decision rules and solutions is possible to determine which are most appropriate in making projections with respect to the behavior of the system.

The major models are compared as to their value in making projections, and the work and money involved. We also discuss many potential policy and administrative changes that could alter the projections. Throughout the body of this report it should be remembered that our concern has been primarily with methodology. Actual projected numbers should be used with caution, particularly those that are set far into the future. Their usefulness lies in the aid they can give forecasters who will make subjective alterations to these numbers based on their experiences.

CHAPTER II

LITERATURE REVIEW

The following is a literature summary of some papers dealing with model fitting and methods of enrollment projections. The first part of this summary concentrates on a number of papers dealing with general regression and projection problems and the second part concentrates on those papers that are concerned with the more specific problems of enrollment projections.

Methods of fitting polynomials to a set of data are discussed in papers by Berztiss (1964), Hager and Antel (1968), and Hudson (1969). The paper by Berztiss compares and discusses a number of methods of fitting curves to irregularly spaced data with regard to their properties and applicability. Methods considered include the use of monomials x^k, Legendre polynomials $P_k(x)$, and Chebyshev polynomials $T_k(x)$. Some remarks are also made on the nature of ill-conditioned matrices, a common problem when parameter estimates are being found for polynomial models. Hager and Antle compare the effectiveness of four methods of determining the correct degree of a polynomial mode: Guttman's, the lack of fit test, the extrapolation test, and backwards regression. The authors concluded that Guttman's method is of no practical value and that any new method for estimating the degree of a polynomial should be compared to the lack of fit test. Hudson's paper provides an algorithm to fit a least squares polynomial subject to a constraint that it be either

non-negative, non-positive, non-decreasing, non-increasing, convex or concave at all points x in a closed interval .

A method for choosing a three-parameter family of curves to fit longitudinal data, the choice being made from a four-parameter family general enough to contain as subfamiles most of the curves frequently fitted, is suggested by Day (1966). The general form of the family of curves considered is

$$f(t) = K\{1 + ae^{-bt+c}\}^{-(1/a)} \quad .$$

A comparison between the growth of individuals, or groups of individuals, now becomes a simple matter of comparing the parameter values.

Of prime interest is a general paper on fitting linear relations by the method of least squares by Anscombe (1967). He ties in the effect of modern computers on statistical calculation with the analysis of data by such methods as vector components, vectors of independent and dependent variables, (instead of analysis of variance), stepwise regression, and testing the goodness of fit by examining the residuals. This paper also includes a discussion, of the problems presented, by a number of statisticians.

Papers by Schatzoff, Tsao, and Fienberg (1968), Valiaho (1969), Hocking and Leslie (1967), and LaMotte and Hocking (1970) consider the problems of stepwise regression and the selection of regression variables. Schatzoff, Tsao, and Fienberg describe efficient computational procedures for calculating all possible

2^k-1 regressions of a dependent variable upon subsets of k independent variables when k is reasonably small as an alternative to the use of stepwise regression procedures. Comparisons of efficiency among five different procedures are also provided. In his paper Valiaho proposes a stepwise method for computing weighted regression models that is stepwise with respect to both observations and variables: one observation can be added or removed from the model at a time in any order.

The papers by Hocking and Leslie, and LaMotte and Hocking deal with the problems of selecting the best subset of independent variables in multiple linear regression analysis. The first paper is concerned with using the C_p - statistic of Mallows as the basic criterion, the second paper extends the basic ideas of the first so that moderately large problems can be treated with a minimum of computation.

Measuring flows of students through an educational system is an integral part of making enrollment projections. Although a variety of methods have been used in the past, [Short (1967), Lins (1960)], Short recommends the "analysis method," presumably determining parameter rates, as the most explicit way of projections. He also recommends doing projections by more than one method to act as checks on each other. Stone (1965) divides education into a number of processes - compulsory and voluntary. He assumes that between each pair of successive processes, transition probabilities are given. A dynamic input-output model is thus formed and changes in the transition probabilities over time are discussed. Stone (1966) elaborates on his approach of a year earlier and develops

a demographic accounting framework as well as a flow model with a transition ratio approach. We adopt his accounting framework in the form of age-grade and cohort tables and attempt to expand on his flow model as a contingency table form.

For many years the two main methods of making enrollment projections were the "grade-cohort" and "enrollment-ratio" using the terminology of Liu (1966). The grade-cohort approach relies on the development of grade-retention ratios which are the ratios of students continuing from one grade to the next. The enrollment-ratio method is based on the projection of past and current ratios of school enrollment and requires estimates of population. Liu provides detailed procedures for both methods and applies them both. His analysis is non-mathematical in nature.

Several authors such as Schmid and Miller (1957), Adair (1962), Jacoby (1959) and Dominion Bureau of Statistics (1965) also examine school systems by studying student progress through schools by means of the cohort survival technique. Schmid and Miller develop retention ratios for each grade which ages cohorts through the system grade by grade. Jacoby makes projections by use of the survival ratios which are closely aligned to some of the parameters we develop. Adair also mentions the "average percentage of income technique" which is only a variation of the two major methods, DBS does not do projections but just examines the flow of students through grades in Canada by means of retention rates. Their other approach is the census method, another variation of the enrollment ratio method.

The cohort survival method is recommended by Davis (1966) as being more applicable than the ratio method. His objective is different from most as he concentrates on determining the enrollment "needed" in developing countries rather than what the enrollment many actually be. Consequently a heavy emphasis is placed on teacher requirements, capital costs and force feeding students. He applies least squares analysis to derive flow rates through an educational system. Enrollment of new students in a grade is expressed as functions of enrollments of total students in previous years.

Brown and Savage (1960) develop models for various prediction situations. They study retrospective longitudinal data, fitting various models to try and predict over the near future. Two methods are used to predict the number of high school graduates. The first uses live births as the independent variable and the number of graduates as the dependent variable related by univariate regression. The authors use the cohort approach with the first grade enrollment as the independent variable. This is also a regression but is less restrictive. Transition ratios form the basis of the second method. These methods are specific cases of the general "grade-cohort" and "enrollment-ratio." Probabilities are estimated for the transition of students as new entrants, processing through the system, and out of the university. Multiple regression models are used to determine the number of students in attendance as a function of high school graduates and net changes in armed services personnel.

Haggstrom (1969), Watson (1969), Simon & Fullam (1969)

are others who fit models in the form of regressions. "Projections" use simple regression methods wherever trends could be established. Their rationale for using the straight line method is that the true curve of a relationship would be likely to yield a straight line over any portion covering only a ten-year span. Ontario uses univariate orthogonal polynomials and fits linear, quadratic and cubic forms. Haggstrom uses hypothetical underlying enrollment rates and assumes that the rate increases by a nonlinear regression model, a logistic growth curve. His parameters are estimated from the data and similiar to Brown and Savage the number of draftees are incorporated into the model. Haggstrom compares his results with USOE Projections and gives as the difference between the two the fact that he allows for returning veterans in its ratios.

Sawiris (1970) suggests several methods that utilize the technique of linear multiple regression analysis to project future college enrollment in one course, a group of courses combined, and a whole college or system. Included are deterministic models with a simple random component, stochastic models, and models containing both deterministic and stochastic components. The author notes that the success of a particular method is a function of the number of observations available, the variance of noise in the data, the serial correlation of noise and other factors. Therefore some experience is needed to develop the skill to select the appropriate method and model that will fit the data.

In recent years there has been a decided emphasis on employing specific Markov models to the field of enrollment projections.

Wurtele (1967) gives a summary of projection models based on transition proportions. Essentially Markov type flow models using mutually exclusive categories. Applications of these models are discussed as well as the sources of change in transition proportions.

Correa (1963) was one of the earlier advocates of using a Markov chain model for an educational system. He studies transitions between three school levels and does not differentiate between specializations. The model does not treat drop-outs at each level. It does consider flows of students as functions of intelligence categories and income levels.

Moser and Redfern (1965) employs another type of Markov model with a large number of processes with transition proportions between each process. The cell and grade-state ideas are introduced here with a cross-classification of sex, age, and educational status.

Gani (1963), Thonstad (1967), Li (1969) and Zabrowski "Dynamod II" (1967), all employ Markov models in the form of Markov chains. Baisuck (1969) discusses whether or not such theory is applicable to the projection of enrollments. His conclusion is that the Markov properties are not all found in the enrollment and flow process and he brings into play a "Markov-Type" model. We will return to his ideas later.

Gani developed formulae for attendance at university using assumptions about fixed transition probabilities between different stages. Dynamod II models students and teachers placing

students into mutually exclusive categories. Another key feature here is projected future net births. The pattern is one of simulation of the real system depicting a country.

Correa and Reimer (1969) construct a simulation model dealing with student flows. The emphasis is made on parameters that can be varied according to the goals of the policy-makers. For each set of values of these parameters one simulation run could be made. All variables were computed internally from enrollment data which was known at the beginning and end of each year. Thonstad follows a deterministic approach pointing out the incompleteness of the application to a Markov chain when only a first order model is used along with a deterministic approach. Transition ratios are assumed to be uniform. Thonstad appears to be the only one to discuss the observed minus expected values. He maintains that any differences are due to non-stationary and migration effects. Li follows the Thonstad model and tries to get into the structure of an educational system. Renewal theories from demography are introduced and an attempt is made to assess the components of the model.

RPI (1968) develops a model for stepping groups of simulated students through an educational system according to the projected parameters. The key feature of the plan is a data information system. Such a system could include subjective estimates, statistical sampling and possibly the development of a data bank on student flow. Costs and benefits of the computer simulation model are discussed. Baisuck bases his work on the RPI

CHAPTER III

DATA BASE

3.1 DATA BASE

Enrollment data for the past 25 years was collected from the Department of Education of Alberta, supplied in an age-grade-sex table for each year. In addition data used in estimating initial parameters such as pass rates, failure rates for grades 1 and 9, dropouts for grades 7 to 12, birth rates, etc., have been collected from a variety of sources such as The Alberta Department of Education, The United States of America Office of Education, The Alberta Bureau of Statistics, The Dominion Bureau of Statistics and The Manitoba Department of Education.

The collected data is tabulated and ordered in a manner dependent on the variables to be used, to give us a data base, from which we may generate mathematical models and make projections of the future student enrollments in Alberta. Projections being defined, as before, as an estimation of a future data base one or more steps along the time interval. We are not attempting primarily to forecast future enrollments but merely provide the theoretical models and projections on which forecasts may be made.

Projections may be based on student transition parameters (rate of student flows) and unless the student movements are very stable we can expect that these parameters will vary with time. In Alberta our system is open, that is, exogenous variables are

included which will effect enrollments and may cause increasing or decreasing trends in the parameter values, quite apart from random variability from year to year. These trends in the transition parameters are often complicated by changes in school administration polices, such as entrance requirements. Thus the data base must be general enough to include these variabilities within its structure and permit adjusted projections to be made using the mathematical models as before.

Students in an educational system can be categorized by many variables, whose selection is dependent upon the goals of the analyses. A student is located in our system by the variables age, grade, sex, and year. Year is either the academic school year or the cohort year, the year in which the student was born.

Data was collected retrospectively over a period of 25 years, with no overlapping of years, consequently we are justified in referring to a year as a discrete variable.

In Alberta the grades are from 1 to 12 and the ages 5 to 21. Age is defined by The Department of Education as the age at the last birthday at the beginning of the accounting period, which in recent years has started from September 30 of each academic year.

Beside age, grade and sex other variables could also be included, e.g., region. However, we limit our variables as we do not wish to over specialize our projection and also if the number of variables considered is large the parameters in the mathematical

models can become so numerous that any computations are made impractical, if not impossible.

Baisuck (1969) refers to the overall education system groupings as a major classification scheme. Our data can be classified in a four-dimensional matrix with the dimensions age, grade, sex and year. However, as an individual can be found many times in this matrix, this is not a contingency table and is not readily suited for computation in this form. It is more pertinent to the problem to classify our information in three-dimensional contingency tables with dimensions age, grade and sex and with one table for each year. Furthermore, since we base many of the analyses on age and grade specific transition probabilities, we can further partition the data into age-grade tables, cross-classification tables showing the variables age and grade for a specific sex in a specified year, or cohort tables, longitudinal tables relating to a group of people born in a particular year. The following section is a more detailed discussion of such tabular representations.

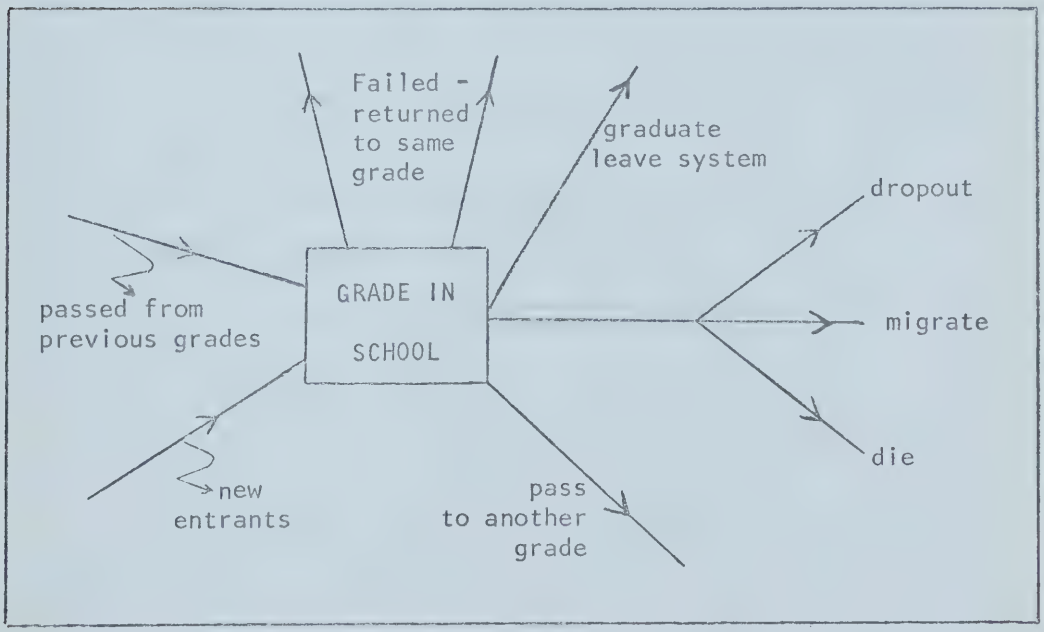
3.1.1 Processes of the System

By tracing the flows of individuals through the system we can show in detail the structure of the student population at any one time. This is the basis upon which we project enrollments and demonstrate how the structure can change under a variety of conditions.

The number of students enrolling in any one grade and

year can be divided into those who have moved up a grade (promoted), those who are repeaters in the grade (failures), and those entering the system for the first time (new entrants). The number of students completing a school year can be divided into those who complete a grade and are promoted, those who complete the grade and fail, and those who drop out of the system (dropouts). This is shown figuratively in Figure 3.1.

FIGURE 3.1

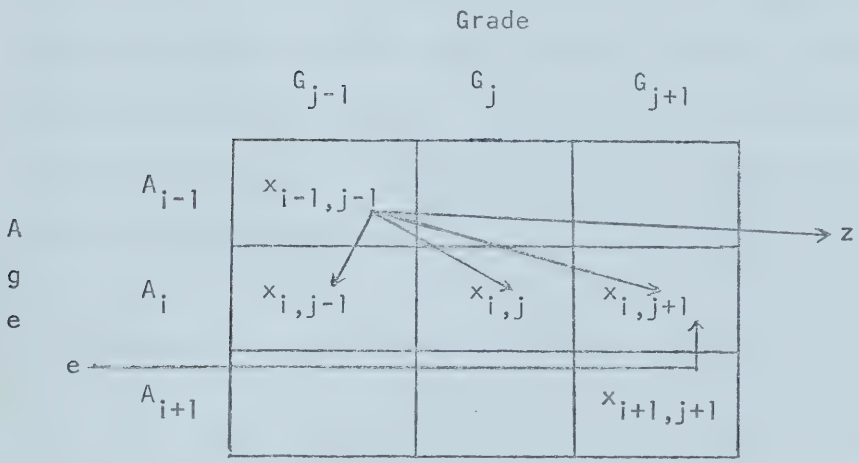


If the analysis is restricted to that closed system there are no exogenous variables, e.g., migration. However, our system is open and in the chapters on a Markov chain model and a random walk simulation inclusion of such parameters as migrants, dropouts, returning dropouts and deaths is discussed, as it is in these models

that these parameters can be included in the projection.

The flow of students can be represented figuratively by looking at a piece of an age-grade table, for a specific sex and for a specified year. This is a table identified by the variables age, grade, sex, and year; any individual cell within the age-grade table can be located if these variables are specified.

FIGURE 3.2



The students found in the cell denoted by $x_{i-1,j-1}$ can flow in a number of directions:

$x_{i-1,j-1} \rightarrow x_{i,j-1}$ = failure in grade $j-1$ and the students repeat that grade the following year

$x_{i-1,j-1} \rightarrow x_{i,j}$ = promotion from grade $j-1$ to grade j

$x_{i-1,j-1} \rightarrow x_{i,j+1}$ = double promotion from grade $j-1$ to grade $j+1$

e = new entrants into the system (at the first grade level this would be individuals now ready for school. At higher levels it would represent individuals entering from outside the system).

z = individuals leaving the system (at the last grade level it may be those graduating. At any level it could either be emigrants or dropouts from the system).

The inclusion of e and z change a closed model into an open model. The flow of students through an educational system is determined by such factors as promotion, failure, double promotion, dropout, mortality, and migration. These rates of flow through the system are discussed as parameters when formulating our mathematical models.

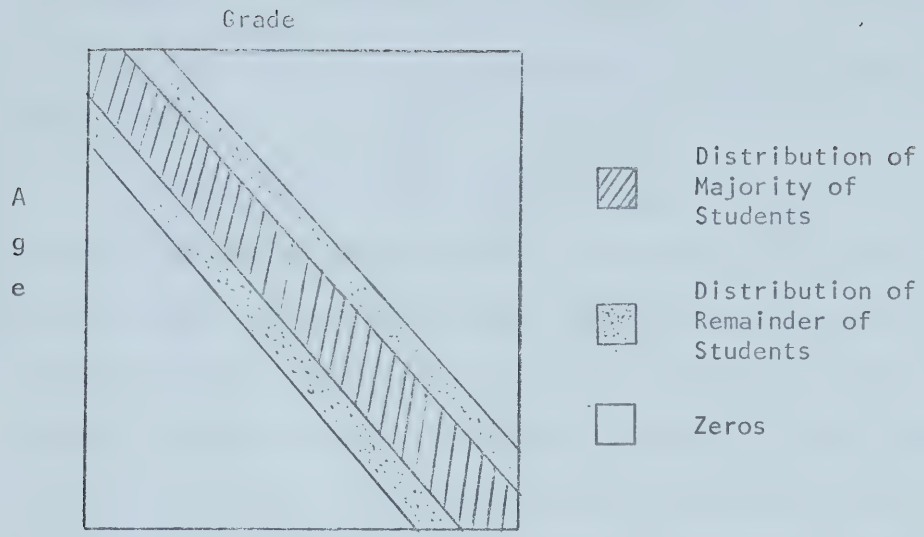
3.1.2 Data Format

A cross sectional table is a contingency table in which variables are measured, at a specific point in time, for each individual who is only to be found once in that table. Age-grade tables are cross sectional tables identified by the variables age, grade, sex, and year, with a table for each sex for each specified year. This cross-sectional data is distributed multinomially since it gives us the distribution of the students by sex and time, amongst the 1 age groups and 12 grades.

An accounting matrix (Stone, 1966) is a numerical statement of the changes that take place in the grade levels of the students between given periods of time. The number of variables

in our study is small compared with the number of individuals, however, regardless of this large number of individuals many cells in the accounting matrix are empty or contain very few observations relative to the other cells. The majority of the individuals are found in a diagonal band running the length of the matrix. This diagonal band usually consists of four or five major diagonals including the main diagonal and its adjacent diagonals. This has been shown figuratively in Figure 3.3. Obviously the younger students will be distributed amongst the lower grades and the older students amongst the higher grades. The main body of students for any age group will usually be in one or two grades, e.g., seven year olds in grades 2 and 3, with a few in higher or lower grades, e.g., seven year olds in grades 1 and 4, thus accounting for the small number in some cells. We will have upper and lower triangular sets of zeros as it is highly improbable to find any student age 7, say, in grades 7 to 12, likewise we will not find any student age 18 in grades 1 to 8. We could classify the matrices as almost being upper triangular zero due to the system being hierarchial. This applies to both the cohort tables and age-grade tables.

FIGURE 3.3



3.2 ACCOUNTING MATRICES

It is necessary to place the data into a consistent framework to provide a base for the development of projection models. We use this base to gain an understanding of the school system and then to test the models.

An accounting matrix is a numerical statement of the changes that take place in the grade level of the students during given periods. The first step in the construction of the matrix is to divide the student population by grade and age, for each sex, for every year that data is available. Other classification variables such as school and region could be included but, as mentioned previously, this increases the number of parameters, necessary to the development of mathematical models, beyond the limits of practicability.

It was stated in the introduction that the aims of the study were projection of grade school enrollments in Alberta and development of a variety of models from which others could make projections of their own school enrollments. Thus we do not include such special variables as region.

Two series of accounting matrices are constructed: (i) age-grade tables, identified by the variables age, grade, sex, and year so that any individual cell can be found within an age-grade table if these variables are specified, and (ii) cohort tables, a series of tables that relate to groups of people born in a particular year. These matrices cover a series of years in which we can observe how the direction and volume of movement changes from year to year. The flow of students into, through and out of the education system are measured and we see how each of these are related to one another. The direction and volume of these flows do not necessarily stay constant through time, a most important factor in determining enrollment projections and the models on which they are based.

3.2.1 Cross-Sectional Tables

These tables are contingency tables with as many dimensions as there are variables where variables are measured on each individual with each individual found only in a given table; each cross-sectional table relates to a single time period at which the measurements are taken on the variables. The main distinction among types of cross-sectional tables is between

those in which age is the prime criterion of classification and those in which all the information is aggregated over age, summary tables. We restrict our attention to the former as they relate to the specific requirements of our problem.

3.2.1.1 Description of Age-Grade Tables

An age-grade table is identified by the variables age, grade, sex, and year. Thus any individual cell within the age-grade table can be found if these variables are specified by looking at a piece of the age-grade table, e.g.,

FIGURE 3.4*

Grade

Age	G_{j-1}	G_j	G_{j+1}
A_{i-1}	$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j-1}$
A_i	$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
A_{i+1}	$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

*This table is measured for a specific sex and for a specified year.

where

G_j = grade level of student, $j = 1,\ldots,12$

A_i = age of student, $i = 5,6,\ldots,20,21$

$x_{i,j}$ = enrollment total of students, age i and grade j.

The marginal totals x_{i+} and x_{+j} refer to the sum of the rows (ages) and columns (grades), respectively.

That is,

$$x_{i+} = \sum_{j} x_{i,j}$$

$$x_{+j} = \sum_{i} x_{i,j}$$

$$N = \sum_{i,j} x_{i,j} \quad ,$$

where N is the total number of observations in an age-grade table for a given sex and year.

3.2.2 Longitudinal Tables

Cohort tables are a special branch of longitudinal tables and relate to the group of people born in a particular year. This will be referred to as a cohort year with specific reference made to the age cohort, the group of pupils entering a school system who are born in the same year. The simplest form of this kind of table arises where the movements of the cohort are recorded year by year. Such tables consist of sub-matrices for the successive years of age taken from a sequence of cross-classification tables in which age is the main basis of classification. In longitudinal studies the position of a

FIGURE 3.5

General Age-Grade Table for the Year y

Age \ Grade	K	1	2	3	...	10	11	12	
5	$x_{5,K,y}$	$x_{5,1,y}$			...				$x_{5.y}$
6	$x_{5,K,y}$	$x_{6,1,y}$	$x_{6,2,y}$		...		(0)		$x_{6.y}$
7		$x_{7,1,y}$	$x_{y,2,y}$	$x_{7,3,y}$	...				.
.					...				.
.					...				.
.					...				.
16					...				.
17		(0)			...	$x_{17,10,y}$	$x_{17,11,y}$	$x_{17,12,y}$	.
18+					...		$x_{18+,11,y}$	$x_{18+,12,y}$	$x_{18+.y}$
	$x_{.Ky}$	$x_{.1y}$			...		$x_{.11y}$	$x_{.12y}$	$x_{..y}$

FIGURE 3.6

Yearly Female Enrollment for 1969-70

AGE	GRADE													TOTAL
	-0-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	
5	0	5479	11	0	0	0	0	0	0	0	0	0	0	5490
6	0	12540	4955	25	0	0	0	0	0	0	0	0	0	17520
7	0	896	12085	4750	57	0	0	0	0	0	0	0	0	17789
8	0	67	1134	11915	4624	119	0	0	0	0	0	0	0	17859
9	0	6	120	1359	11695	4799	182	0	0	0	0	0	0	18161
10	0	0	18	136	1436	10972	4737	90	0	0	0	0	0	17389
11	0	0	0	30	214	1451	10428	4386	131	0	0	0	0	16649
12	0	0	0	9	35	295	1593	10407	4018	170	0	0	0	16527
13	0	0	0	0	12	62	325	1732	9779	4234	156	0	0	16300
14	0	0	0	0	0	21	61	362	1886	9503	3924	233	0	15990
15	0	0	0	0	0	14	18	75	479	1845	8835	3698	162	15106
16	0	0	0	0	0	0	0	15	70	393	1648	7534	3316	12976
17	0	0	0	0	0	0	0	0	13	79	383	1357	7174	9006
18	0	0	0	0	0	0	0	0	0	20	62	274	1739	2095
19	0	0	0	0	0	0	0	0	0	0	18	54	237	409
20	0	0	0	0	0	0	0	0	0	0	7	15	83	110
21	0	0	0	0	0	0	0	0	0	8	7	14	112	141
TOTAL	0	19988	18323	18224	18073	17732	17344	17067	16376	16252	15100	13179	13028	202506

cohort is often only recorded periodically. The only difference between cross-sectional tables and cohort tables is that for cohort tables the enrollments are for a single cohort over a sequence of years rather than for a set of cohorts in a single year.

3.2.2.1 Description of Cohort Tables

Each cohort table is identified by the variables age, grade, sex and year. Thus any individual cell within the cohort table can be found if these variables are specified. By looking at a piece of a cohort table

FIGURE 3.7

Age \ Grade	G_{j-1}	G_j	G_{j+1}
A_{i-1}	$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j+1}$
A_i	$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
A_{i+1}	$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

we can trace the flows of students for any given sex and year, where

A_i = age of students, $i = 5,6,\ldots$,

G_j = grade of students, $j = 1,2,\ldots,12$

$x_{i,j}$ = enrollment of students, age i and grade j.

If we used sex as a third variable we would have a three-dimensional matrix as follows:

FIGURE 3.8

S_1

Age \ Grade	G_{j-1}	G_j	G_{j+1}
A_{i-1}	$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j+1}$
A_i	$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
A_{i+1}	$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

FIGURE 3.9

S_2

Age \ Grade	G_{j-1}	G_j	G_{j+1}
A_{i-1}	$x_{i-1,j-1}$	$x_{i-1,j}$	$x_{i-1,j+1}$
A_i	$x_{i,j-1}$	$x_{i,j}$	$x_{i,j+1}$
A_{i+1}	$x_{i+1,j-1}$	$x_{i+1,j}$	$x_{i+1,j+1}$

where S_k = sex of students, $k = 1,2$.

Cohort year could be employed as the fourth variable giving us a four dimensional matrix with

$$y_\ell = \text{cohort year}, \qquad \ell = 1,2,\ldots \ .$$

A cohort table has structure different from a common contingency table because it may contain an individual student as many as twelve times. Students forming a cohort are aged through the educational process, i.e., they may be found initially in a cell with enrollment $x_{i-1,j-1}$, $x_{i-1,j}$ or $x_{i-1,j+1}$; then in a cell with enrollment $x_{i,j-1}$, $x_{i,j}$ or $x_{i,j+1}$ and finally in a cell denoted by enrollment $x_{i+1,j-1}$, $x_{i+1,j}$ or $x_{i+1,j+1}$, depending on how they advance through the system. Thus, unlike the age-grade table, each student will be found in every row since each successive year the individual is one year older although he does not necessarily move ahead a grade level. Cohort analysis is that where a group of students is followed through the system from entrance to final status, dropout, graduation, death or migration.

3.2.3 Relationship of Age-Grade Tables to Cohort Year Tables

When we inspect slices of multivariate longitudinal data for which the time variable is held constant we are dealing with cross-sectional data. As an example consider the enrollment data from a regular school system. In any year enrollment data may be collected by age, grade, and sex, this is cross-sectional data that forms a contingency table by placing the number of

individuals, identical with respect to all variables, in the same category. A complication of this data over a period of years yield longitudinal data in the form of cohort tables. Thus any individual can appear in a cohort table anywhere from zero to 12 times, the total number of grade-states. This is shown in Figure 3.12 where cohorts are represented as staircases.

The stairs shown represent the cohort that is 18 years old in 1965. Other cohorts are represented by staircases parallel to the one shown. An age-grade cohort table is represented by looking at the front face of the staircase (i.e., the projection of the staircase on the front plane).

Age-grade tables and cohort year tables are related to each other as a function of year and age.

$$\text{cohort year} = (\text{age-grade year}) - \text{age} \quad .$$

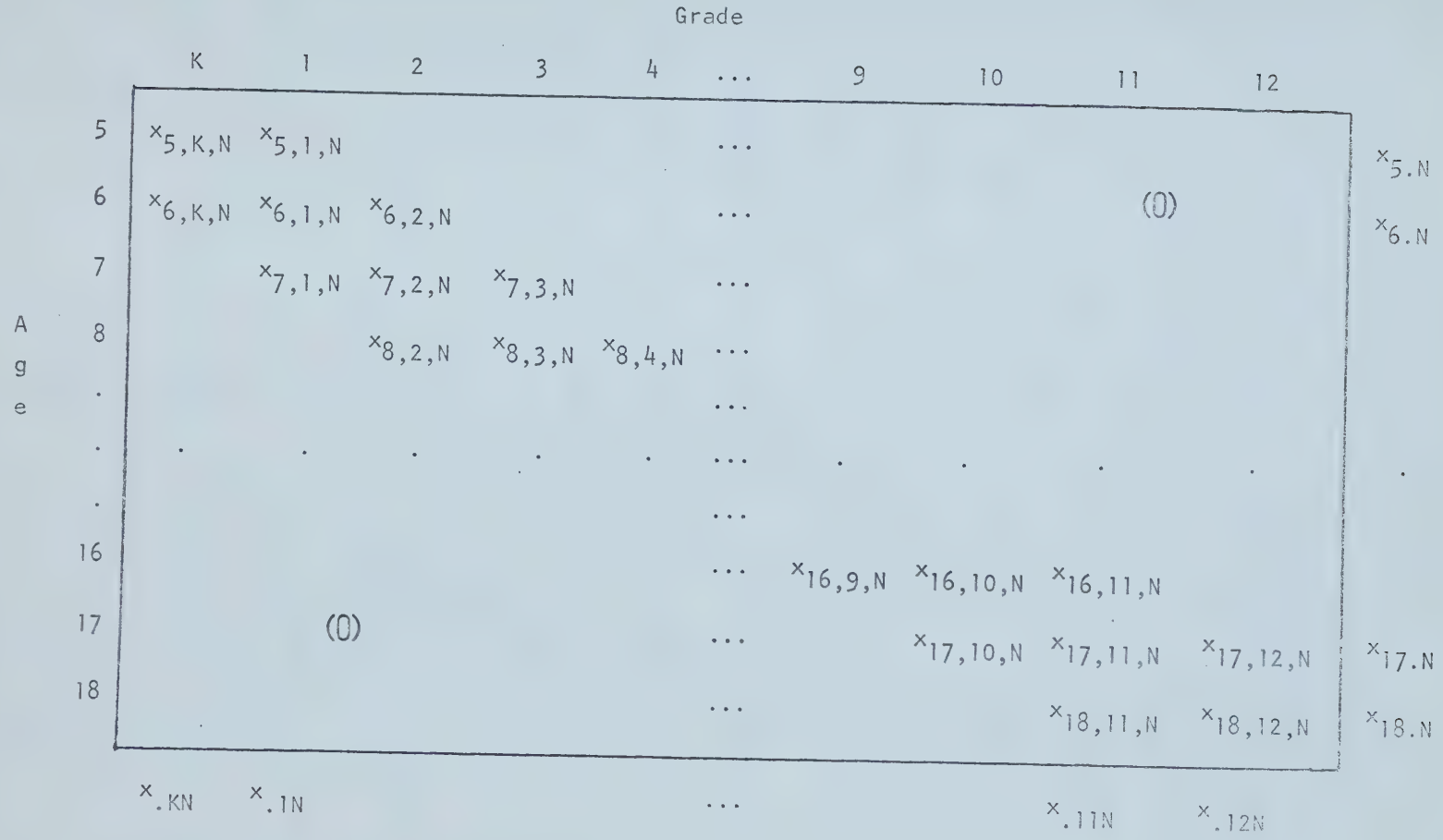
This means that we can go from the enrollment at a certain age for a given age-grade year and find this same enrollment in the cohort table once we have calculated the cohort year.

3.3 FLOW MATRICES

In an educational system students move from lower to higher grades as they fulfill the academic requirements of the lower grade. These movements are defined as flows through the system. Proportional representation of these student flows can be made by forming transition matrices. These transition matrices contain elements which give the relative frequencies by which

FIGURE 3.10

General Cohort Table for Those Students Age 18 in Year N



Age 18 in Year N

FIGURE 3.11

COHORT TABLE FOR MALES, 10 IN 58-59

AGE	GRADE													TOTAL
	-0-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	
5	0	973	1	0	0	0	0	0	0	0	0	0	0	974
6	0	6009	1545	7	0	0	0	0	0	0	0	0	0	7561
7	0	2131	3698	1756	81	0	0	0	0	0	0	0	0	7665
8	0	258	2412	3135	2118	0	0	0	0	0	0	0	0	7923
9	0	89	419	2728	2496	1954	228	0	0	0	0	0	0	7903
10	0	51	136	695	2559	2520	2151	77	0	0	0	0	0	8189
11	0	23	62	220	849	2272	2543	2083	23	0	0	0	0	8125
12	0	14	32	107	259	1213	1892	2575	1681	100	0	0	0	7873
13	0	2	22	69	105	269	1162	1621	2673	1219	0	0	0	7147
14	0	0	10	31	41	95	309	1469	1913	2675	1418	2	0	7963
15	0	0	5	16	22	51	83	427	1230	1401	2160	921	35	6421
16	0	0	0	0	6	6	22	104	344	1210	976	2272	607	5549
17	0	0	0	0	0	0	2	28	55	389	858	1145	2413	4890
18	0	0	0	0	0	0	0	6	8	73	258	491	1559	2395
19	0	0	0	0	0	0	0	0	0	10	41	125	449	625
20	0	0	0	0	0	0	0	0	0	1	9	24	145	179
21	0	0	0	0	0	0	0	0	0	2	0	14	73	89
TOTAL	0	9550	8341	8764	8525	8382	8392	8340	7932	7160	5720	4794	5281	91181

AGE	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
MEAN GRADE	1.00	1.21	1.97	2.90	3.82	4.72	5.64	6.51	7.38	8.35	9.21	10.17	11.10	11.46	11.62	11.75	11.76
VARIANCE	0.00	0.17	0.55	0.69	0.99	1.13	1.28	1.56	1.60	1.60	1.83	1.61	1.18	0.74	0.46	0.32	0.31

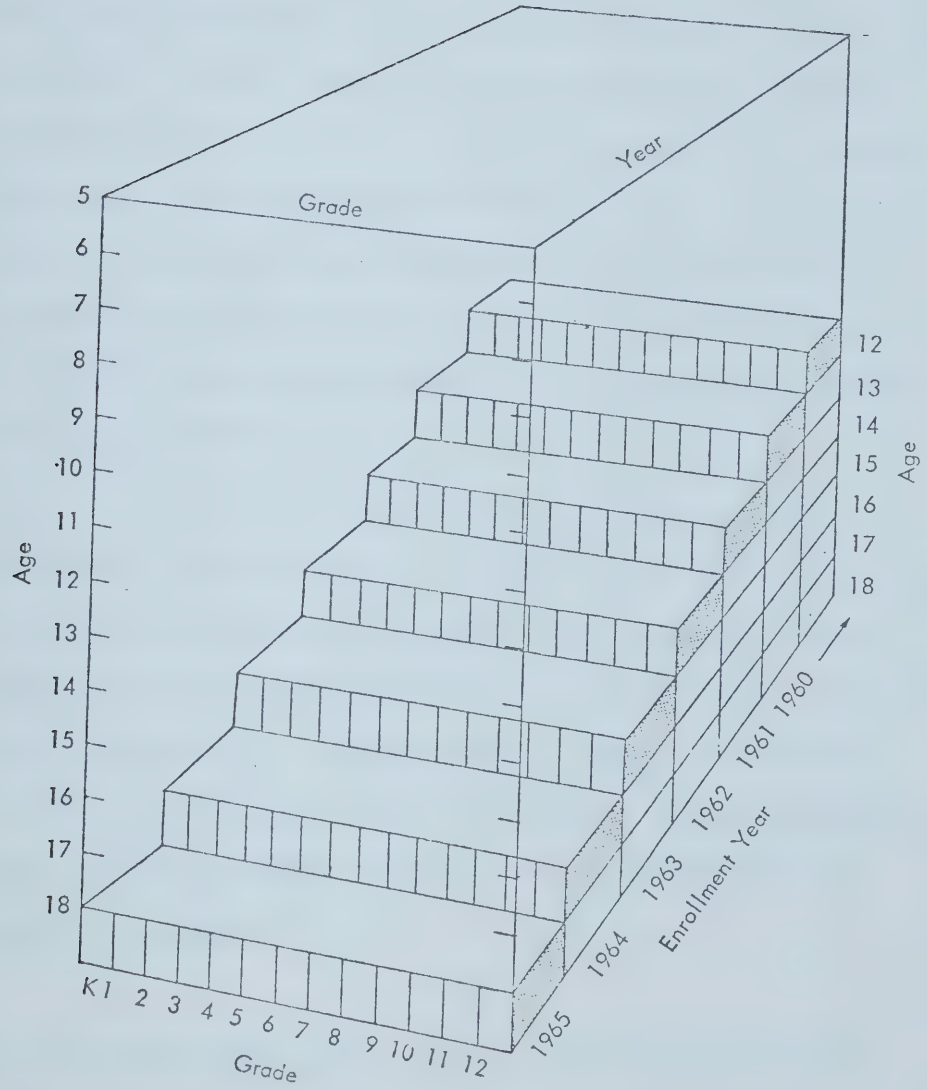
COHORT TABLE FOR FEMALES, 10 IN 58-59

AGE	GRADE													TOTAL
	-0-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	
5	0	772	0	0	0	0	0	0	0	0	0	0	0	772
6	0	6120	1170	14	0	0	0	0	0	0	0	0	0	7304
7	0	1429	4688	1201	0	0	0	0	0	0	0	0	0	7318
8	0	157	1613	4087	1095	37	0	0	0	0	0	0	0	7994
9	0	63	212	1866	3424	1823	235	0	0	0	0	0	0	7623
10	0	25	70	352	1676	3567	2251	145	0	0	0	0	0	8086
11	0	12	33	109	354	1666	2997	2453	0	0	0	0	0	7624
12	0	5	21	64	124	516	1627	2970	1955	131	0	0	0	7413
13	0	0	11	30	63	130	557	1398	3129	1605	0	0	0	6923
14	0	0	3	26	41	49	138	760	1720	3152	1682	0	0	7571
15	0	0	0	1	24	24	51	172	700	1522	2813	1403	0	6710
16	0	0	0	0	6	18	21	47	135	827	1017	2477	824	5572
17	0	0	0	0	0	2	0	21	24	157	490	728	2218	3630
18	0	0	0	0	0	0	0	0	0	22	106	171	791	1090
19	0	0	0	0	0	0	0	0	0	0	13	13	89	115
20	0	0	0	0	0	0	0	0	0	0	0	1	19	20
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	0	8563	7831	7750	7707	7832	7877	7966	7663	7416	6111	4993	3941	85670

AGE	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
MEAN GRADE	1.00	1.16	1.97	3.02	3.98	4.98	5.94	6.84	7.75	8.66	9.63	10.52	11.35	11.59	11.66	0.0	11.95
VARIANCE	0.0	0.14	0.36	0.56	0.79	0.86	0.97	1.19	1.17	1.29	1.29	1.31	0.97	0.56	0.45	0.0	0.05

FIGURE 3.12

Cohort Tables



students move among the grades of the educational system.

A system of students and student flows can be arranged into the form of a standard matrix, the marginal vectors of the matrix give the accounting figures for two successive years or two consecutive rows of a cohort table. The entries in the matrix show the flows connecting these accounting figures. These matrices are based on year to year flows and consist of row and columns with the sets placed in a different time relationship to each other. The accounting matrix refers to one time period, usually a year, while the flow matrix shows the movement of students from one time period to another time period. The difficulty arises in determining how specific students move from one year to the next.

A flow matrix is an input-output matrix, that is, a system of connected processes with the outputs being shown in the rows and the inputs in the columns. These processes are not completely interdependent but form a one way hierarchy of dependence with the lower processes always feeding the higher ones. In our system the input consists of young children entering the school system and output consists of those who dropout, die, migrate or graduate.

3.3.1 Contingency Tables Derived from the Multidimensional Data

As an example consider a set of cohort tables based on student enrollment data collected over a 15 year period. By considering each sex separately there appear to be two ways of

looking at this body of data that will yield three-dimensional contingency tables. One way is to arrange the data by successive year pairs. The first structure will be referred to as "year wise" and the second as "age wise". Each arrangement yields a set of three-dimensional contingency tables with a table corresponding to each year or age pair.

3.3.1.1 Year Wise

The layers of each contingency table are different age pairs for the same year pair. The enrollment data covers the period 1946 to 1970 hence our year pairs are (46-47, 47-48), (47-48, 48-49), ..., (68-69, 69-70). Each year pair defines a three dimensional contingency table composed of a series of two-dimensional contingency tables, with one layer for each successive age pair. In our example the age pairs are 5-6, 6-7, ..., 17-18. Thus for the year pair 53-54, 54-55 we have a contingency table for each of these age pairs which together form a three-dimensional contingency table. Two of these tables are illustrated in Figure 3.13. In the first table the row totals give the 1953-54 enrollment distribution for nine year old girls. The second table gives the distribution for the same years but for the ages ten and eleven.

3.3.1.2 Age Wise

In this case the layers of each contingency table are the different year pairs for the same age pairs. Thus the age wise

FIGURE 3.13

Year Wise Tables

Age 10 Females (1954-55)

Age 9 Females (1953-54)

Grade	Drop Outs	1	2	3	4	5	6	7	8	9	Total
New Entrants											
1											2924
2											7353
3											11257
4											6701
5											975
6											53
7											1
8											
9											
Total		1130	2925	7347	10133	6215	1116	75	5	0	

Age 11 Females (1954-55)

Age 10 Females (1953-54)

Grade	Drop Outs	1	2	3	4	5	6	7	8	9	Total
New Entrants											
1											1469
2											3730
3											7068
4											10205
5											5214
6											915
7											75
8											1
9											
Total		501	1344	3377	6500	9267	5298	1050	96	3	

tables are formed in the same manner as the year wise tables with the roles of the year and age pairs reversed. Two layers from a contingency table for the age wise structure are given in Figure 3.14. As can be seen, the difference between Figures 3.13 and 3.14 is that in the former the year pairs remain constant whereas in the latter it is the age pairs that are constant. It is clear that each two-dimensional contingency table constituting a layer in the year wise structure is also contained in the year wise structure. Only the arrangement of the layers is changed.

The three basic variables involved are age, grade, and year. If we let:

(i) a and a' denote age where $a' = a+1$ and $a = 5,6,\ldots,17$

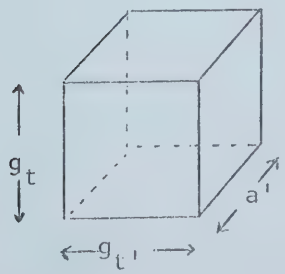
(ii) g and g' denote grade where $a' = g+1$ and $g = 1,2,\ldots,11$

and

(iii) t and t' denote year where $t' = t+1$ and $t = 46\text{-}47,\ 47\text{-}48,\ \ldots,\ 68\text{-}69$

then the year wise and age wise cohort table takes the form

YEAR WISE



AGE WISE

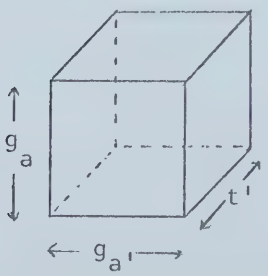


FIGURE 3.14

Age Wise Tables

Age 11 Females (1954-55)

Grade	Drop Outs	1	2	3	4	5	6	7	8	Total
New Entrants										
1										1469
2										3730
3										7068
4										10205
5										5214
6										915
7										74
8										1
9										
Total		503	1344	3377	6500	9267	5298	1050	96	

Age 10 Females (1953-54)

Age 11 Females (1955-56)

Grade	Drop Outs	1	2	3	4	5	6	7	8	Total
New Entrants										
1										1130
2										2925
3										7347
4										10133
5										6215
6										1116
7										75
8										5
9										
Total		491	1314	4458	8278	9046	4896	1115	61	

Age 10 Females (1954-55)

for the year wise table the grade and time variables combine to produce the three variables; age, grade in year t, and grade in year t', and for the age wise table the age and grade variables; grade at age a, grade at age a', and year. In general the grade variable will also include a category for new entrants when it is a column variable and a category for drop-outs when it is a row variable as well as cells for grades 1 to 12. This is shown in Figures 3.13 and 3.14.

Usually all the row and column totals, the enrollment totals for each age, grade, and year, are known as well as the cell totals for given age, grade, and year combinations, but not the cell totals for the grade by grade combinations. To estimate values for these unknown totals we must make use of available information and make some assumptions. It is in such an instance that the general Markov model can be applied.

3.4 MODELS

A model of students moving through an education system uses various sets of assumptions in making projections. These assumptions pertain to the distribution of the students in the system and the transition proportions tracing their flows. We want to investigate and eliminate many models to find those that best describe student flows and totals and consequently, use the same or other models to project these flows and totals into the future.

We employ both stochastic and deterministic models and both open and closed models. The closed models are not affected by external variables such as mortality and migration, the open models take into account students entering or leaving the system. Open models, such as Markov chain and random walk simulation are affected by the exogenous variable parameters, closed models, such as time series are only dependent on the present state of the system.

The mathematical details and the merits of each model are discussed in detail in the following chapters. Chapter 4 uses a variety of models to discover trends and tendencies in the data base. Chapters 5, 6, and 7 actually implement and extend some of the findings from Chapter 4 to make enrollment projections.

CHAPTER IV

DATA ANALYSIS

4.1 INTRODUCTION

The initial stages of this project involved analysis of the data by fitting a variety of models to the Alberta enrollment data to investigate the nature of the data and to detect trends and patterns in student enrollment that exist in the data. This was done prior to the actual selection and application of the models that were eventually used to project future student enrollment for the system. The results obtained from these initial fittings permitted a greater understanding of the nature of the data at hand that could be obtained by simply viewing the enrollment tables in their original form. These results were helpful in deciding which models should be discarded as not being appropriate and which models might be retained for further application, as well as in suggesting additional models and assumptions that could be placed on the models.

The models applied may be categorized as follows: (1) polynomial models, (2) auto-regressive models, and (3) the giant regression model. The polynomial models include four types: polynomial regression on cohort year, on age, and on grade, as well as polynomial regression on year, age, and grade simultaneously. The auto-regressive models consist of two and three parameter auto-regression, two cell averages on year and

on age-grade combination, and auto-regression through cohort year and along cohort diagonal.

In addition to the model fittings the yearly grade distribution for each age and sex was tabulated by enrollment and by percentage of enrollment for each grade, and the yearly age distribution was tabulated for each grade and sex by enrollment and percentage of enrollment for each age. Descriptions of the models used are given along with examples of the output for some of the models.

4.2 POLYNOMIAL MODELS

Four different polynomial models were fitted to the Alberta enrollment data. These were (i) polynomial regression on cohort year, (ii) polynomial regression on grade, (iii) polynomial regression on age, and (iv) polynomial regression on age, grade, and year.

For the first three cases the polynomial model has the general form

$$x = \beta_0 + \beta_1 z + \beta_2 z^2 + \ldots + \beta_m z^m + e \quad , \qquad (4.1)$$

where x is the dependent variable to be estimated, z is the independent variable or predictor, e is the model error, and m is the order of the polynomial model. As applied in the present situation the model essentially consists of fitting a curve to a set of data points by the method of least squares and estimating the dependent variable x by

$$\hat{x} = b_0 + b_1 z + b_2 z^2 + \ldots + b_m z^m \quad , \qquad (4.2)$$

where the $\{b_i\}$ are the least squares estimates of the parameters $\{\beta_i\}$.

In the fourth case the model has the form,

$$x_{ijn} = \beta_0 + \beta_1 + \beta_2 j + \beta_3 n + \beta_4 ij + \beta_5 in + \beta_6 jn + \beta_7 i^2 \qquad (4.3)$$
$$+ \beta_8 j^2 + \beta_9 n^2 + \beta_{10} i^2 j + \beta_{11} i^2 n + \beta_{12} ij^2 + \ldots + e_{ijn} \, .$$

In this case the model assumes some structure about the relationship of x_{ijn} to the variables i, j, n, and can be thought of as representing a summation of polynomials in i, j, and n. Once again estimates of the dependent variable x_{ijn} were obtained by the method of least squares.

4.2.1 Polynomial Regression on Cohort Year

These are polynomials fitted to the same age-grade cell in a set of student cohort tables. The polynomials were fitted to the cohort tables for each age-grade combination. This produced a set of polynomials that gave estimated enrollment totals for each combination of age, grade and year. The polynomial model has the form,

$$x_{ijn} = \beta_0 + \beta_1 n + \beta_2 n^2 + \ldots + \beta_m n^m + e_{ijn} \quad , \qquad (4.4)$$

where the dependent variable x_{ijn} is the student enrollment total for age i, grade j, and cohort year n, and the independent

variable is cohort year n.

4.2.2 Polynomial Regression on Grade

In this case the polynomials were fitted to the cohort tables for fixed cohort year n and age i, using grade j as the independent variable or predictor. Thus a polynomial fit was obtained for each cohort year-age combination in the set of cohort tables. The model is

$$x_{ijn} = \beta_0 + \beta_1 j + \beta_2 j^2 + \ldots + \beta_m j^m + e_{ijn} . \qquad (4.5)$$

Each cohort table has a corresponding set of polynomials consisting of one polynomial for each age.

4.2.3 Polynomial Regression on Age

This approach is similar to that described in 4.2.2 except in this case cohort year n and grade j are kept fixed and age becomes the predictor. Here, the model is

$$x_{ijn} = \beta_0 + \beta_1 i + \beta_2 i^2 + \ldots + \beta_m i^m + e_{ijn} . \qquad (4.6)$$

4.2.4 Polynomial Regression on Age, Grade, and Year

In this case each of the three variables age i, grade j, and cohort year n, as well as combinations of these variables are simultaneously used as predictors for the student enrollment total x_{ijn}. In this sense each of the three previous polynomial models are combined to give a single predictive model

incorporating as independent variables the three variables that have previously been taken separately. The model is

$$x_{ijn} = \beta_0 + \beta_1 i + \beta_2 j + \beta_3 n + \beta_4 ij + \beta_5 in + \beta_6 jn + \beta_7 i^2 + \beta_8 j^2 + \ldots + e_{ijn} \quad . \tag{4.7}$$

4.2.5 Discussion and Comparison

The basic objective of polynomial fitting is to obtain a polynomial that fits the data adequately and is of low order. By employing polynomials of sufficiently high order it is possible to produce curves that fit the data exactly. Of course, such a procedure is of little value for our purpose as each enrollment total would be estimated exactly and would therefore hide any trends that exist within the data. With this in mind the four approaches to fitting polynomials discussed above were carried out for comparison with each other and the other methods to be described.

Polynomials from the first to the fifth order were fitted to the enrollment data in the first three cases. The usual approach taken to obtain an adequate polynomial model is to first fit the linear or first order polynomial

$$x = \beta_0 + \beta_1 z + e \quad .$$

An F-test of significance of the highest degree term is then made. If it proves significantly different from zero the second order polynomial is then fitted,

$$x = \beta_0 + \beta_1 z + \beta_2 z^2 + e \quad ,$$

and once again the significance of the highest degree term is tested.

This procedure is continued to higher order polynomials until the highest degree term is found non-significant, and then the previously fitted polynomial, one less in order, is taken as the model for the data. If after a number of fittings the highest degree term is still found to be significant, it is possible the polynomial model is inappropriate and a different type of model must then be considered.

In the present study polynomials from the first to the fifth order were fitted simultaneously to the data by computer and the probability of the highest degree term tabulated for each case.

Other measures of fit that were also computed for each polynomial are the multiple correlation coefficient R^2 and the Durbin-Watson statistic D. The formula for the multiple correlation coefficient is

$$R^2 = \frac{\sum (\bar{x} - \hat{x})^2}{\sum (\bar{x} - x)^2} \quad ,$$

where x and $\hat{x}$ are the observed and fitted values and $\bar{x}$ is the mean of the observations. Thus, R^2 is the proportion of the variance of x accounted for by the model. The range of R^2 is zero to one with values near one indicating a good fit.

The significance of its value, however, depends on the number of observations, therefore, direct comparisons of R^2 values can only be meaningful for polynomial models fitted to the same number of data points. This is an important consideration when trying to determine the effect of increasing the order of the polynomial fitted.

The Durbin-Watson statistic is given by

$$D = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2},$$

where the $\{e_t\}$ are the calculated residuals. That is, $e_t = x_t - \hat{x}_t$, where x_t is the observed value and $\hat{x}_t$ is its least squares estimate. The range of D is 0 to 4, with values of D near 0 or 4 indicating auto-correlation of the residuals. Auto-correlation can be due to the inadequacy of the model or the existence of a time dependent error structure. The method of least squares assumes independence of the residuals and if the assumption is not met an under-estimation of the true residual variance usually results and with it a distortion of the F-values, R^2-values, and standard errors, producing an exaggerated appearance of precision in the regression.

Figures 4.1 to 4.6 are examples of the summary tables obtained for the polynomial regression. The tables presented here are summaries of the number of observations, the square of the multiple correlation coefficient, the probability of the

highest degree term equal to zero, and the Durbin-Watson statistic for the cubic polynomial regression on cohort year.

The best fits, as indicated by the high correlations in Figure 4.2 occur at each grade for the age group with the largest student enrollment in the grade. For example, the polynomial with the highest correlation for grade six students occurs at what might be termed the 'usual' age for this grade, this is age eleven, and the square of the multiple correlation coefficient is 0.995876. This result might be expected, due to the smaller variance in enrollment that occurs for these age-grade combinations from year to year.

Figure 4.3 gives the probability of the highest degree term, the cubic term in this case, being equal to zero. The stars indicate those cubic terms that are significantly different from zero and the five per cent level. Their percentage of the total is 42.6. In these cases a quartic polynomial would be fitted and the highest degree term tested for significance in the manner previously described. Figure 4.4 gives the Durbin-Watson statistic calculated for each polynomial regression. Values less than one or greater than three tend to suggest the model is inappropriate. The true significance of the statistic, which depends on the number of observations, may be obtained by reference to a table of critical values.

Figures 4.5 and 4.6 illustrate the detailed description of cubic polynomial regression on cohort year for the age six, grade one combination. For this case R^2 is 0.989 and the

FIGURE 4.1

MALES CUBIC TIME SERIES REGRESSION THROUGH COHORTS

NUMBER OF OBSERVATIONS

AGE	GRADE					
	1	2	3	4	5	6
5	25.	21.	0.	0.	0.	0.
6	25.	25.	24.	0.	0.	0.
7	25.	25.	25.	25.	0.	0.
8	25.	25.	25.	25.	8.	0.
9	25.	25.	25.	25.	25.	25.
10	24.	25.	25.	25.	25.	25.
11	23.	25.	25.	25.	25.	25.
12	23.	25.	25.	25.	25.	25.
13	21.	23.	24.	25.	25.	25.
14	5.	22.	23.	25.	25.	25.
15	0.	20.	23.	23.	25.	25.
16	0.	0.	0.	21.	24.	25.
17	0.	0.	0.	0.	0.	21.
18	0.	0.	0.	0.	0.	0.
19	0.	0.	0.	0.	0.	0.
20	0.	0.	0.	0.	0.	0.
21	0.	0.	0.	0.	0.	0.

	7	8	9	10	11	12
5	0.	0.	0.	0.	0.	0.
6	0.	0.	0.	0.	0.	0.
7	0.	0.	0.	0.	0.	0.
8	0.	0.	0.	0.	0.	0.
9	0.	0.	0.	0.	0.	0.
10	25.	0.	0.	0.	0.	0.
11	25.	18.	0.	0.	0.	0.
12	25.	25.	25.	1.	0.	0.
13	25.	25.	25.	13.	0.	0.
14	25.	25.	25.	25.	17.	0.
15	25.	25.	25.	25.	25.	17.
16	25.	25.	25.	25.	25.	25.
17	24.	25.	25.	25.	25.	25.
18	21.	23.	25.	25.	25.	25.
19	0.	0.	25.	25.	25.	25.
20	0.	0.	18.	24.	25.	25.
21	0.	0.	21.	1.	25.	25.

FIGURE 4.2

MALES CUBIC TIME SERIES REGRESSION THROUGH COHORTS

SQUARE OF THE MULTIPLE CORRELATION COEFFICIENT

AGE	GRADE					
	1	2	3	4	5	6
5	0.904907	0.595345	0.000000	0.000000	0.000000	0.000000
6	0.888246	0.862637	0.693917	0.000000	0.000000	0.000000
7	0.965798	0.988154	0.953054	0.164501	0.000000	0.000000
8	0.990990	0.981610	0.989810	0.868596	0.890295	0.000000
9	0.740359	0.941342	0.487950	0.989066	0.925228	0.377976
10	0.985723	0.944732	0.980034	0.117423	0.992355	0.953774
11	0.194244	0.899213	0.981292	0.981673	0.689349	0.995876
12	0.954613	0.951526	0.977474	0.997787	0.999174	0.930203
13	0.998427	0.852377	0.914986	0.932660	0.978781	0.982107
14	0.000000	0.544809	0.894138	0.952621	0.964943	0.962303
15	0.000000	0.143248	0.896823	0.773213	0.904706	0.968004
16	0.000000	0.000000	0.000000	0.231721	0.053066	0.848519
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.747814
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	7	8	9	10	11	12
5	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.114821	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.973397	0.821563	0.000000	0.000000	0.000000	0.000000
12	0.996114	0.974760	0.388593	0.000000	0.000000	0.000000
13	0.954103	0.992835	0.972039	0.884773	0.000000	0.000000
14	0.956339	0.948595	0.977625	0.941938	0.237434	0.000000
15	0.960506	0.930038	0.986079	0.965946	0.554383	0.395168
16	0.952936	0.960430	0.992451	0.979530	0.880946	0.840066
17	0.921532	0.738525	0.923149	0.704131	0.984429	0.985162
18	0.241315	0.640831	0.812417	0.912022	0.912814	0.965541
19	0.000000	0.000000	0.416474	0.302285	0.592932	0.967078
20	0.000000	0.000000	0.876766	0.088751	0.695328	0.957154
21	0.000000	0.000000	0.754395	0.000000	0.705638	0.971955

FIGURE 4.3

MALES CUBIC TIME SERIES REGRESSION THROUGH COHORTS

PROBABILITY OF THE HIGHEST DEGREE TERM

AGE	GRADE					
	1	2	3	4	5	6
5	0.77701	0.58862	[illegible]	0.00000	0.00000	0.00000
6	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
7	0.22841	0.01952*	0.09437	0.97466	0.00000	0.00000
8	0.00011*	0.11617	0.04186*	0.58945	0.13412	0.00000
9	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
10	0.00146*	[illegible]	0.00002*	0.41951	0.00947*	0.81392
11	0.23723	0.00014*	0.00000*	0.00009*	0.12290	0.00071*
12	0.17157	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
13	0.00003*	0.05855	0.00205*	0.00030*	0.00700*	0.41672
14	0.00005	0.54006	0.20163	0.00231*	0.06471	0.67059
15	0.20000	0.15059	0.04617*	0.51942	0.04186*	0.71742
16	0.00000	0.00000	0.00000	0.08203	0.80530	0.58451
17	0.00000	0.00000	0.00000	0.00000	0.00000	0.00253*
18	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
21	[illegible]	[illegible]	0.00000	0.00000	0.00000	0.00000

	7	8	9	10	11	12
5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
6	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
8	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
9	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
10	0.29867	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
11	0.40732	0.76977	0.00000	0.00000	0.00000	0.00000
12	0.00007*	0.91874	0.25655	0.00000	0.00000	0.00000
13	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	0.81007	0.19486	0.00051*	0.00139*	0.66759	0.00000
15	0.87293	0.29141	0.00001*	0.06253	0.01975*	0.83887
16	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
17	0.40989	0.25727	0.00131*	0.50082	0.00000*	0.11942
18	0.06407	0.54624	0.01261*	0.06744	0.03933*	0.00844*
19	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
20	0.00000	0.00000	0.00648*	0.54308	0.74711	0.11918
21	0.00000	0.00000	0.04829*	0.00000	0.17925	0.27849

46 SIGNIFICANT OUT OF 108 REGRESSIONS, OR 42.6%

FIGURE 4.4

MALES CUBIC TIME SERIES REGRESSION THROUGH COHORTS

DURBIN-WATSON STATISTIC

AGE	GRADE					
	1	2	3	4	5	6
5	1.9085	2.7233	0.0000	0.0000	0.0000	0.0000
6	1.9814	2.1934	2.0977	0.0000	0.0000	0.0000
7	2.2149	1.3992	1.7058	2.3526	0.0000	0.0000
8	1.4161	1.0651	2.4296	2.4309	2.1853	0.0000
9	2.3741	1.5829	2.3314	2.0200	1.5947	2.1123
10	1.9859	1.7110	1.5647	2.2321	1.7644	1.9081
11	1.6754	2.3539	1.3908	1.7356	2.0734	1.2042
12	1.8753	1.9124	1.9785	1.7092	1.4603	1.8739
13	1.9567	2.2455	2.2136	2.3660	0.9181	0.8828
14	0.0000	2.5021	1.5972	3.0794	1.1192	1.2166
15	0.0000	1.4554	2.9255	0.8609	2.0291	1.4316
16	0.0000	0.0000	0.0000	2.2678	3.4608	3.3470
17	0.0000	0.0000	0.0000	0.0000	0.0000	1.9539
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

	7	8	9	10	11	12
5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	2.3019	1.4017	0.0000	0.0000	0.0000	0.0000
11	2.3819	1.4017	0.0000	0.0000	0.0000	0.0000
12	1.4911	2.2766	2.5034	0.0000	0.0000	0.0000
13	1.1561	1.3105	1.9567	2.7662	0.0000	0.0000
14	1.1683	1.7702	1.4261	1.2850	2.1160	0.0000
15	1.1013	1.5136	2.1401	1.5016	1.7262	2.2848
16	2.4269	1.7115	1.8960	1.4165	1.7912	1.4871
17	2.2543	2.0686	1.7330	2.2973	1.7803	1.4121
18	1.4688	3.2501	2.1379	2.8048	2.1899	1.5151
19	0.0000	0.0000	2.6545	3.0575	2.4254	1.4515
20	0.0000	0.0000	2.5747	1.7803	2.7286	1.8054
21	0.0000	0.0000	2.4838	0.0000	2.3468	1.4197

FIGURE 4.5

CUBIC TIME SERIES REGRESSION THROUGH COHORT YEARS

MALES AGE 6 GRADE 1

NUMBER OF OBSERVATIONS 25

STANDARD ERROR OF REGRESSION 0.3019897E 03

R-SQUARED 0.9992462

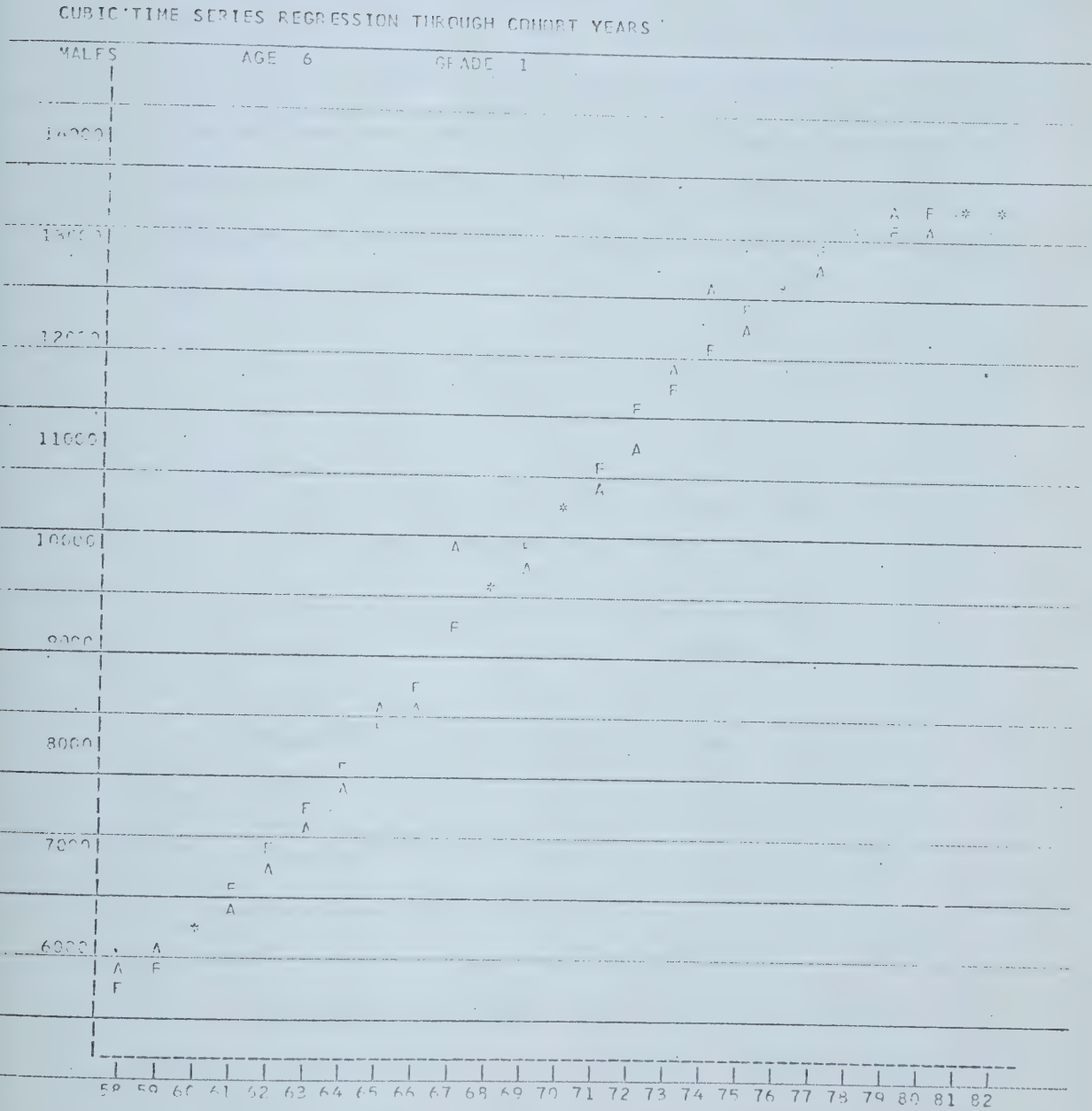
SOURCE	DF	SS	MS	F	P
REGRES	3.	0.1761767E 09	0.5872557E 08	643.93604	0.0000010
RESID	21.	0.1915155E 07	0.9119781E 05		
TOTAL	24.	0.1780918E 09			

VAR.NO.	WEIGHT	STND.ERROR	T-VALUE	P
1	10561.3[illegible]685	91.7174143	116.4197997	0.0000007
2	442.3[illegible]86	21.0579071	20.9119568	0.0000006
3	-6.[illegible]87	1.3017282	-5.3202858	0.0000297
4	-[illegible]	0.2768582	-3.9566569	0.0007217

COHORT	ACTUAL	FITTED	RESIDUAL	PERCENT
58	5853.	5694.	159.	2.72
59	[illegible]	[illegible]	45.	0.67
60	6221.	6284.	-63.	-1.01
61	6[illegible]	[illegible]	-195.	-2.[illegible]
62	6[illegible]	7[illegible]	-99.	-1.24
63	7[illegible]	7[illegible]	-68.	-0.93
64	7[illegible]	7[illegible]	-205.	-2.68
65	8[illegible]	8[illegible]	116.	1.36
66	8[illegible]	8741.	-233.	-2.71
67	1[illegible]	9[illegible]	962.	9.47
68	9777.	9659.	118.	1.[illegible]
69	9828.	10115.	-287.	-2.92
70	10436.	10561.	-125.	-1.20
71	1[illegible]	1[illegible]	-296.	-2.77
72	11[illegible]	11408.	-2[illegible]	-2.[illegible]4
73	11973.	11798.	175.	1.44
74	12781.	12160.	631.	4.94
75	12314.	12488.	-174.	-1.41
76	12610.	12777.	-167.	-1.3[illegible]
77	12657.	12[illegible]4.	-167.	-1.3
78	13271.	13222.	48.	0.36
79	13440.	13367.	73.	0.[illegible]4
80	13305.	13454.	-149.	-1.[illegible]
81	13[illegible].	13478.	12.	0.09
82	13531.	13434.	97.	0.72

DURBIN-WATSON STATISTIC 1.[illegible]

FIGURE 4.6


CUBIC TIME SERIES REGRESSION THROUGH COHORT YEARS
MALES AGE 6 GRADE 1
11000
10000
9000
8000
7000
6000
58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82

Durbin-Watson statistic is 1.98, which indicate an excellent fit. The polynomial equation is

$$x_{1,6,n} = 1056.33 + 440.36k + 6.93k^2 - 0.82k^3 \quad , \qquad (4.8)$$

where $k = n-\bar{n}$. This transformation of cohort year was performed in each case to improve the ill-conditioned nature of the normal equations, and hence the accuracy of the parameter estimates that can result when higher order polynomials are fitted. Figure 4.6 gives plots of the actual and estimated enrollment totals obtained from equation (4.8).

FIGURE 4.7

Percentage of Significant Highest Degree Terms
for Polynomial Regression on Cohort Year

MODEL	MALES	FEMALES
Linear	89	85
Quadratic	65	66
Cubic	43	38
Quartic	17	11
Quintic	15	12

Figure 4.7 indicates that a third order polynomial will provide an adequate fit to the data for most age-grade combinations, the inclusion of higher order terms being necessary in only a small percentage of cases. These latter cases were found in most instances

to correspond to age-grade combinations with small numbers of students, for example, grade one - age twelve.

The polynomial regression on age and on grade were found to provide poor fits to the data and only high order polynomials produced large R^2 values even though the number of observations were considerably less in these cases. The poor fits are due to the greater variance of enrollment totals for these cases, which would require a different type of model to adequately describe the data.

Figures 4.8, 4.9 and 4.10 describe the fitting of the polynomial regression on age, grade, and year for males to the Alberta enrollment data. Stepwise regression has been used on the variables listed in Figure 4.8. A, G and Y correspond to age, grade, and year, and A3 and Y2G refer to A^3 and Y^2G, etc.

Figure 4.9 gives a description of the computed estimation equation, including the value of R^2, each of the variables in the equation, and their order of inclusion with their estimated parameter estimates. It can be seen that the best predictors of enrollment are age, grade, and year, and age, grade, and year squared in that order. Figure 4.10 lists each variable included in the equation, in order, with the value of R^2 for the equation at each stage.

As indicated by the low value of $R^2 = 0.334$ prediction is quite poor using this method. However, it does serve to

FIGURE 4.8

GIANT POLYNOMIAL

FILE ALBERTA (CREATION DATE = 01/23/71) REVISED DATA

VARIABLE	MEAN	STANDARD DEV	CASES
X	1367.3568	2277.3047	2455
Y	58.0289	7.1913	2455
A	13.2037	3.7948	2455
G	6.2163	3.3733	2455
Y2	3419.0497	835.2938	2455
A2	189.7512	100.5321	2455
G2	50.9167	44.4818	2455
YA	766.1185	241.6555	2455
YG	361.4138	203.7891	2455
AG	91.9808	66.4904	2455
Y2A	45134.0859	17345.3844	2455
Y2G	21336.2823	13138.5822	2455
A2Y	10952.2945	6055.8404	2455
A2G	1433.1976	1305.2601	2455
G2Y	2911.8285	2643.4987	2455
Y3	204359.0269	73650.7158	2455
A3	2868.2627	2167.2293	2455
G3	456.8933	525.7168	2455
YAG	5342.2778	3967.6287	2455
G2A	796.4859	814.9491	2455
A4	45722.0346	44217.9012	2455
G4	4479.4978	6120.5258	2455
Y4	12387332.1515	5839669.7669	2455
A3G	23311.1340	25590.6672	2455
A3Y	166507.4473	129144.9363	2455
G3A	7455.2383	9521.5719	2455
G3Y	26623.2037	31121.5550	2455
Y3A	2697851.0570	1276363.7156	2455
Y3G	1278073.9857	879901.5521	2455
A2G2	12852.3527	15346.7766	2455
A2Y2	645311.2240	389832.5743	2455
G2Y2	172133.0660	164156.4579	2455
Y2AG	315279.5267	249097.0111	2455
A2YG	83340.1279	77506.7506	2455
G2YA	45798.1206	49283.8576	2455

FIGURE 4.9

GIANT POLYNOMIAL

FILE ALBERTA (CREATION DATE = 01/23/71) REVISED DATA

* M U L T I P L E R E G R E S S I O N *

DEPENDENT VARIABLE.. X

VARIABLE(S) ENTERED ON STEP NUMBER 14.. G3Y

MULTIPLE R	0.57778	ANALYSIS OF VARIANCE	DF	SUM OF SQUARES	MEAN SQUARE	F
R SQUARE	0.33383	REGRESSION	14.	4246176586.40736	303441184.74338	87.33731
STANDARD ERROR	1863.96331	RESIDUAL	2440.	8477436493.01586	3474359.21845	

VARIABLES IN THE EQUATION

VARIABLE	B	BETA	STD ERROR B	F
A	1454.05292	2.42305	476.50581	9.312
G	-1027.12617	-1.58160	414.08407	6.153
Y	442.53721	1.39752	132.92688	11.082
A2	-27.93425	-1.22322	38.04672	0.539
G2	-137.91781	-2.69206	31.13561	19.803
Y2	-1.13298	-0.41555	0.96187	1.387
AG	137.37315	4.01106	28.69270	22.922
YA	-31.41409	-3.33364	4.13361	57.755
YG	4.19616	0.37552	6.33649	0.439
A3	-0.44630	-0.42475	1.05122	0.180
G3	8.47630	1.95684	3.00016	7.982
Y2AG	0.01558	1.70443	0.00519	9.027
A2G2	-0.15020	-1.01225	0.06163	5.940
G3Y	-0.06939	-0.95518	0.03741	3.491
(CONSTANT)	-14874.08335			

VARIABLES NOT IN THE EQUATION

VARIABLE	BETA IN	PARTIAL	TOLERANCE	F
Y2A	0.13471	0.00155	0.00009	0.006
Y2G	1.10556	0.01483	0.00012	0.536
A2Y	-4.02477	-0.04678	0.00009	5.348
A2G	-93.26586	-0.45407	0.00002	640.553
G2Y	0.43716	0.00263	0.00002	0.016
Y3	-5.88566	-0.02837	0.00002	1.964
YAG	-1.19476	-0.00890	0.00004	0.193
G2A	18.98510	0.10794	0.00002	28.431
A4	4.53533	0.03507	0.00004	3.004
G4	0.90262	0.01571	0.00020	0.502
Y4	-2.01414	-0.02644	0.00013	1.974
A3G	-52.98561	-0.42256	0.00004	530.176
A3Y	-2.22440	-0.04811	0.00031	5.653
G3A	33.61945	0.25698	0.00004	172.457
Y3A	-0.05983	-0.00163	0.00050	0.007
Y3G	0.38111	0.01335	0.00082	0.435
A2Y2	-3.85963	-0.06291	0.00018	9.691
G2Y2	0.78164	0.01287	0.00018	0.404
A2YG	-10.50984	-0.16847	0.00017	71.247
G2YA	4.85513	0.05008	0.00007	6.132

F-LEVEL OR TOLERANCE-LEVEL INSUFFICIENT FOR FURTHER COMPUTATION

FIGURE 4.10

GIANT POLYNOMIAL

FILE ALBERTA (CREATION DATE = 01/23/71) REVISED DATA

* M U L T I P L E R E G R E S S I O N *

DEPENDENT VARIABLE.. X

SUMMARY TABLE

VARIABLE	MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
A	0.38138	0.14545	0.14545	-0.38138	1454.05392	2.42305
G	0.47614	0.22671	0.08126	-0.10931	-1027.18617	-1.52160
Y	0.50968	0.25977	0.03306	0.19625	442.53721	1.[illegible]
A2	0.51746	0.26777	0.00799	-0.35686	-27.[illegible]	-1.[illegible]
G2	0.51893	0.26929	0.00152	-0.10628	-137.[illegible]	-2.[illegible]
Y2	0.51921	0.26958	0.00029	0.13751	-[illegible]	-0.41[illegible]
AG	0.54966	0.29774	0.02816	-0.18538	137.37315	4.01106
YA	0.55419	0.30713	0.00938	-0.29218	-31.41403	-3.33364
YG	0.56925	0.32404	0.01691	-0.06253	4.19714	0.[illegible]
A3	0.57418	0.32968	0.00564	-0.33086	-0.44630	-0.42475
G3	0.57424	0.32975	0.00008	-0.10248	6.47630	1.95684
Y2AG	0.57562	0.33134	0.00159	-0.11421	0.21952	1.70443
A2G2	0.57695	0.33288	0.00153	-0.16753	-0.15020	-1.01225
G3Y	0.57778	0.33383	0.00095	-0.08027	-0.06989	-0.95518
(CONSTANT)					-14874.08335	

illustrate the relative importance of each variable in predicting student enrollment. An estimate of an enrollment total for a given age A, grade G, and year Y using the equation would be given by

$$x_{AGY} = b_0 + b_1A + b_2G + b_3Y + b_4A^2 + \ldots + \quad (4.9)$$
$$+ b_{13}A^2G^2 + b_{14}G^3Y \quad ,$$

where, b_0 is the constant and $b_1, b_2, \ldots, b_{14}$ are the parameter estimates corresponding to the particular independent or predictor variables as given in Figure 4.9. Prediction of known enrollment totals and projections of enrollment were made using equation (4.9), but as mentioned above the results were quite poor with negative totals appearing in many instances.

4.3 AUTO-REGRESSIVE MODELS

4.3.1 Two- and Three-Parameter Regression

By these methods estimates of the enrollment totals for the cohort tables were obtained by fitting least squares regression equations through cohort with neighbouring cell totals as the independent variables. The dependent variables, enrollments for given age-grade combination through cohort years, are expressed as a function of the number of students who were in the previous grade and the number who were in the same grade the year before. The rationale behind this procedure is that the number of students in a particular age-grade combination is mainly composed of students

who failed the grade or passed into the grade the year before.

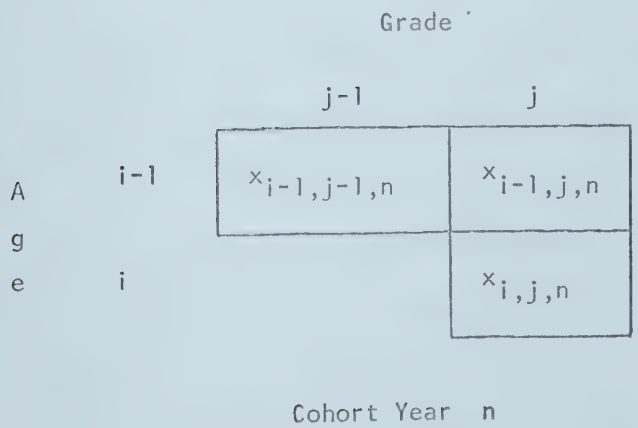
Two-Parameter Regression

The two-parameter regression equation is

$$x_{ijn} = \beta_1 x_{i-1,j-1,n} + \beta_2 x_{i-1,j,n} + e_{ijn} \quad , \qquad (4.10)$$

where the dependent variable is the enrollment total x_{ijn}, with the subscripts i, j, and n specifying the enrollment total by age, grade and cohort year, respectively. The dependent variables $x_{i-1,j-1,n}$ and $x_{i-1,j,n}$ are the particular enrollment totals indicated by the subscripts, β_1 and β_2 are the unknown parameters to be estimated, and e_{ijn} is the model error. The relationship between these variables in the cohort table is illustrated by the following diagram.

FIGURE 4.11



Three-Parameter Regression

The equation for the three-parameter regression has the same form as the two-parameter regression equation with the addition of a constant term. The equation is:

$$x_{ijn} = \beta_0 + \beta_1 x_{i-1,j-1,n} + \beta_2 x_{i-1,j,n} + e_{ijn} \quad , \qquad (4.11)$$

where the variables x_{ijn}, $x_{i-1,j-1,n}$, and $x_{i-1,j,n}$ are as defined for the two-parameter case and β_0 is the constant term.

The fittings for each equation are summarized by tabulating the number of observations for each regression, the least squares estimates of the unknown parameters with their t values and probabilities, and the standard error of regression.

The standard error of regression is

$$SE = \sqrt{\frac{\sum (x - \hat{x})^2}{N - r}} \quad ,$$

where x and $\hat{x}$ are the observed and fitted values and N and r are the number of observations and the number of parameters estimated, respectively.

In addition to the summary more complete descriptions of each two- and three-parameter regression were also obtained. These include the tables of actual totals and residuals with the residuals as percentages of the actual totals.

Examples of the actual output are given in Figures 4.12 to 4.17. The estimated parameters are tabulated in Figures 4.12 and

FIGURE 4.12

TWO PARAMETER REGRESSION FOR ALL COHORT YEARS FOR MALES

AGE	GRADE											
	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-
6	0.0	0.867	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	32.466	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.544	0.916	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	1.191	1.894	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	1.045	0.842	0.996	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.056	0.278	2.021	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	1.128	1.066	0.943	0.984	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.043	0.028	-0.026	-0.337	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	1.134	1.610	0.834	0.948	0.953	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.080	0.001	0.257	0.093	1.277	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	1.065	0.775	1.022	0.822	0.961	0.945	0.0	0.0	0.0	0.0	0.0
	0.0	0.041	0.161	0.053	0.059	0.034	0.616	0.0	0.0	0.0	0.0	0.0
12	0.0	-0.720	0.975	1.023	1.195	0.723	0.954	0.792	0.0	0.0	0.0	0.0
	0.0	0.758	0.193	0.203	0.051	0.082	0.101	4.016	0.0	0.0	0.0	0.0
13	0.0	0.868	0.609	0.664	0.997	0.952	0.659	0.970	0.930	0.0	0.0	0.0
	0.0	0.234	0.428	0.131	0.012	0.014	0.128	-0.026	-1.056	0.0	0.0	0.0
14	0.0	0.394	-0.043	0.461	0.423	0.896	1.084	1.201	0.912	0.821	0.0	0.0
	0.0	0.416	0.445	0.000	0.181	0.054	0.078	-0.056	0.231	3.881	0.0	0.0
15	0.0	0.0	-0.845	0.645	0.851	0.693	0.370	0.761	0.440	1.076	0.333	0.0
	0.0	0.0	0.788	0.053	0.130	0.047	0.207	0.072	0.201	-0.483	-0.352	0.0
16	0.0	0.0	0.0	-0.579	0.445	0.273	0.053	0.730	0.904	0.482	0.860	0.893
	0.0	0.0	0.0	0.681	-0.018	0.074	0.216	0.022	0.056	0.165	2.232	-1.234
17	0.0	0.0	0.0	0.0	0.0	0.430	-0.402	-0.380	0.784	0.704	1.056	0.761
	0.0	0.0	0.0	0.0	0.0	-0.061	0.329	0.257	0.063	0.004	-0.048	0.952
18	0.0	0.0	0.0	0.0	0.0	0.0	-0.576	0.096	-0.420	0.433	0.369	0.827
	0.0	0.0	0.0	0.0	0.0	0.0	0.244	0.087	0.237	0.103	0.163	0.262
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.239	0.415	0.102	-0.293
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.116	0.059	0.212	0.367
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.139	-0.226	-0.177
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.157	0.248	0.297
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.736	0.826
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.624	0.510

FIGURE 4.13

THREE PARAMETER REGRESSION FOR ALL COHORT YEARS FOR MALES

AGE	GRADE -1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-
6	0.0	865.942	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.616	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	51.549	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	-2892.801	149.992	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	1.138	0.862	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.131	4.120	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	1393.857	-896.601	1166.702	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.490	0.839	0.742	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	-0.010	0.582	-1.704	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	-53.750	3083.562	-1746.003	620.708	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	1.013	-0.101	0.819	0.842	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.090	-0.062	0.860	-0.820	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.586	-42.384	2368.162	130.956	831.736	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	1.127	1.586	0.036	0.962	0.805	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.080	0.021	-0.005	0.011	-1.276	0.0	0.0	0.0	0.0	0.0	0.0
11	0.0	-3.150	-57.918	-4.246	1574.057	-910.408	-38.580	0.0	0.0	0.0	0.0	0.0
	0.0	1.239	1.511	1.022	0.206	0.874	0.956	0.0	0.0	0.0	0.0	0.0
	0.0	0.039	0.099	0.055	0.048	0.561	0.693	0.0	0.0	0.0	0.0	0.0
12	0.0	2.820	-19.816	-22.994	1117.887	900.047	-713.613	-427.590	0.0	0.0	0.0	0.0
	0.0	-0.743	1.362	1.089	0.991	0.297	0.876	1.009	0.0	0.0	0.0	0.0
	0.0	0.719	0.168	0.038	-0.357	0.101	0.550	1.950	0.0	0.0	0.0	0.0
13	0.0	-7.381	-53.886	-2.159	-40.066	125.094	122.757	-285.210	-677.474	0.0	0.0	0.0
	0.0	0.333	2.562	0.695	0.762	0.912	0.579	0.929	1.131	0.0	0.0	0.0
	0.0	0.536	0.245	0.127	0.099	-0.025	0.137	0.197	1.405	0.0	0.0	0.0
14	0.0	0.082	3.218	1.149	-3.817	-26.293	1015.987	710.144	96.587	259.329	0.0	0.0
	0.0	0.387	-0.287	0.452	0.491	0.713	0.612	0.504	0.928	0.737	0.0	0.0
	0.0	0.412	0.472	0.086	0.169	0.121	-0.172	0.093	0.141	3.569	0.0	0.0
15	0.0	0.0	1.023	3.122	-1.941	-1.616	-158.646	462.801	-500.672	-1332.093	+691.733	0.0
	0.0	0.0	-0.940	0.568	0.902	0.762	-0.080	0.571	0.859	0.915	1.277	0.0
	0.0	0.0	0.785	0.033	0.128	0.031	0.419	-0.024	0.114	0.671	-0.164	0.0
16	0.0	0.0	0.0	3.293	11.381	3.989	-8.214	-10.756	264.519	-194.147	28.344	-644.399
	0.0	0.0	0.0	-0.489	-0.454	0.165	-0.143	0.702	0.740	0.726	0.866	1.222
	0.0	0.0	0.0	0.469	0.141	0.091	0.275	0.040	0.018	0.098	0.299	0.320
17	0.0	0.0	0.0	0.0	0.0	2.793	0.989	3.193	-95.347	664.891	279.980	695.532
	0.0	0.0	0.0	0.0	0.0	0.270	-0.421	-0.355	0.679	0.235	0.687	0.781
	0.0	0.0	0.0	0.0	0.0	-0.123	0.324	0.241	0.174	-0.066	0.033	0.339
18	0.0	0.0	0.0	0.0	0.0	0.0	-0.815	8.803	13.247	-6.333	392.601	-1.711
	0.0	0.0	0.0	0.0	0.0	0.0	-0.697	-0.025	-0.529	0.429	-0.011	0.889
	0.0	0.0	0.0	0.0	0.0	0.0	0.282	-0.012	0.217	0.112	0.114	0.262
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.424	33.589	41.669	-300.543
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.143	0.304	0.227	0.476
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.034	-0.045	0.169	0.312
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.231	11.993	-102.781
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.125	-0.397	0.686
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.118	0.213	0.265
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-14.441	19.298
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.010	-0.166
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.974	0.555

FIGURE 4.14

ESTIMATED CELL TOTALS Y=B1X1 + B2X2, THE RESIDUALS AND THEIR PERCENTAGES OF ACTUAL

MALES, COHORT YEAR 58-59

AGE	GRADE											
	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-
6	0	876	0	0	0	0	0	0	0	0	0	0
	0	669	0	0	0	0	0	0	0	0	0	0
	0.0	43.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0	5109	1429	0	0	0	0	0	0	0	0	0
	0	-1411	327	0	0	0	0	0	0	0	0	0
	0.0	-38.2	18.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0	2433	3601	1990	0	0	0	0	0	0	0	0
	0	-21	-466	128	0	0	0	0	0	0	0	0
	0.0	-0.9	-14.9	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0	395	2657	2916	0	0	0	0	0	0	0	0
	0	23	71	-430	0	0	0	0	0	0	0	0
	0.0	5.5	2.6	-17.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0	134	675	2415	2536	2153	0	0	0	0	0	0
	0	2	20	143	-16	-2	0	0	0	0	0	0
	0.0	1.5	2.9	5.6	-0.6	-0.1	0.0	0.0	0.0	0.0	0.0	0.0
11	0	59	217	846	2252	2495	2079	0	0	0	0	0
	0	3	3	3	20	48	-46	0	0	0	0	0
	0.0	4.8	1.4	0.4	0.9	1.9	-2.3	0.0	0.0	0.0	0.0	0.0
12	0	30	100	248	1130	1849	2631	1703	0	0	0	0
	0	2	7	11	83	43	-56	-22	0	0	0	0
	0.0	6.3	6.5	4.2	6.8	2.3	-2.2	-1.3	0.0	0.0	0.0	0.0
13	0	21	65	105	272	1182	1576	2454	1457	0	0	0
	0	1	4	0	-3	-20	45	224	-238	0	0	0
	0.0	4.5	5.8	0.0	-1.1	-1.7	2.8	8.4	-19.5	0.0	0.0	0.0
14	0	9	29	41	92	303	1386	1795	2724	0	0	0
	0	1	2	0	3	6	83	118	-49	0	0	0
	0.0	10.0	6.5	0.0	3.2	1.9	5.7	6.2	-1.8	0.0	0.0	0.0
15	0	0	15	22	47	80	413	1255	1380	2195	1321	0
	0	0	1	0	4	3	9	-25	101	-35	-400	0
	0.0	0.0	6.3	0.0	7.8	3.6	2.1	-2.0	6.8	-1.6	-43.4	0.0
16	0	0	0	5	8	20	96	338	1193	1071	2165	779
	0	0	0	1	0	2	8	6	17	-95	-93	-172
	0.0	0.0	0.0	16.7	0.0	9.1	7.7	1.7	1.4	-9.7	-4.5	-28.3
17	0	0	0	0	0	2	25	48	351	855	931	2155
	0	0	0	0	0	0	3	7	38	3	214	258
	0.0	0.0	0.0	0.0	0.0	0.0	10.7	12.7	9.8	0.3	18.7	10.7
18	0	0	0	0	0	0	5	7	68	257	502	1578
	0	0	0	0	0	0	1	1	5	1	-11	-19
	0.0	0.0	0.0	0.0	0.0	0.0	16.7	12.5	6.8	0.4	-2.2	-1.2
19	0	0	0	0	0	0	0	0	10	44	131	428
	0	0	0	0	0	0	0	0	0	-3	-6	21
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-7.3	-4.8	4.7
20	0	0	0	0	0	0	0	0	0	7	21	111
	0	0	0	0	0	0	0	0	0	2	3	34
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	22.2	12.5	23.4
21	0	0	0	0	0	0	0	0	0	0	8	93
	0	0	0	0	0	0	0	0	0	0	6	-20
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	42.9	-27.4

FIGURE 4.15

ESTIMATED CELL TOTALS Y=A + B1X1 + B2X2 THE RESIDUALS AND THEIR PERCENTAGES OF ACTUAL

MALES, COHORT YEAR 58-59

AGE	GRADE -1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-
6	0	1516	0	0	0	0	0	0	0	0	0	0
	0	29	0	0	0	0	0	0	0	0	0	0
	0.0	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0	4146	1511	0	0	0	0	0	0	0	0	0
	0	-448	245	0	0	0	0	0	0	0	0	0
	0.0	-12.1	14.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0	2399	3226	2331	0	0	0	0	0	0	0	0
	0	13	-91	-213	0	0	0	0	0	0	0	0
	0.0	0.5	-2.9	-10.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0	401	2647	2643	0	0	0	0	0	0	0	0
	0	17	81	-157	0	0	0	0	0	0	0	0
	0.0	4.1	3.0	-6.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0	114	678	2453	2544	2113	0	0	0	0	0	0
	0	2	17	105	-24	38	0	0	0	0	0	0
	0.0	1.5	2.4	4.1	-1.0	1.8	0.0	0.0	0.0	0.0	0.0	0.0
11	0	60	216	846	2223	2498	2070	0	0	0	0	0
	0	2	4	3	50	45	-37	0	0	0	0	0
	0.0	3.2	1.8	0.4	2.2	1.8	-1.8	0.0	0.0	0.0	0.0	0.0
12	0	30	101	243	1148	1631	2631	1669	0	0	0	0
	0	2	6	11	65	61	-56	12	0	0	0	0
	0.0	6.3	5.6	4.2	5.4	3.2	-2.2	0.7	0.0	0.0	0.0	0.0
13	0	21	67	105	277	1184	1571	2438	1363	0	0	0
	0	1	2	0	-8	-22	50	240	-144	0	0	0
	0.0	4.5	2.9	0.0	-3.0	-1.9	3.1	9.0	-11.8	0.0	0.0	0.0
14	0	9	29	41	93	306	1447	1775	2754	0	0	0
	0	1	2	0	2	3	22	138	-79	0	0	0
	0.0	10.0	6.5	0.0	2.1	1.0	1.5	7.2	-3.0	0.0	0.0	0.0
15	0	0	15	22	47	80	432	1255	1446	2066	1110	0
	0	0	1	0	4	3	-5	-25	35	94	-197	0
	0.0	0.0	6.3	0.0	7.8	3.6	-1.2	-2.0	2.4	4.4	-21.4	0.0
16	0	0	0	5	8	19	97	338	1200	1092	2175	491
	0	0	0	1	0	3	7	6	10	-116	-103	116
	0.0	0.0	0.0	16.7	0.0	13.6	6.7	1.7	0.8	-11.9	-5.0	19.1
17	0	0	0	0	0	2	25	49	348	885	1017	2518
	0	0	0	0	0	0	3	6	41	-27	128	-105
	0.0	0.0	0.0	0.0	0.0	0.0	10.7	10.9	10.5	-3.1	11.2	-4.4
18	0	0	0	0	0	0	5	7	68	256	513	1578
	0	0	0	0	0	0	1	1	5	2	-22	-19
	0.0	0.0	0.0	0.0	0.0	0.0	16.7	12.5	6.8	0.8	-4.5	-1.2
19	0	0	0	0	0	0	0	0	9	44	131	419
	0	0	0	0	0	0	0	0	1	-3	-6	30
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	-7.3	-4.8	6.7
20	0	0	0	0	0	0	0	0	0	7	22	101
	0	0	0	0	0	0	0	0	0	2	2	44
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	22.2	8.3	30.3
21	0	0	0	0	0	0	0	0	0	0	9	95
	0	0	0	0	0	0	0	0	0	0	5	-22
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	35.7	-30.1

4.13 for the two- and three-parameter cases, and the estimated enrollments, and their errors of estimate with their per cent of the actual enrollment for cohort year 1958-59 are given for each model in Figures 4.14 and 4.15. The R^2 values for each three-parameter regression are tabulated in Figure 4.16. Validity of R^2 requires the inclusion of a constant term in the model a result these values were not computed for the two-parameter case.

To illustrate the method of estimation for the three-parameter regression, we would estimate the age 6, grade 2 male enrollment total for cohort year 1958-59 as

$$\hat{x}_{6,2,59} = 865.942 + 0.616x_{5,1,59} + 51.549x_{5,2,59}$$
$$= 1516 .$$

The actual enrollment total is 1545.

For the two-parameter regression the estimates b_1 and b_2 of the unknown parameters β_1 and β_2 can be considered to give some kind of rough estimates of the proportion of students, in the same age group, who pass into the estimated grade from the previous grade and the proportion of students who fail the estimated grade. However, since parameter estimates with values less than zero and greater than one can occur it is not always possible to give them a meaningful interpretation in this sense. A glance at the parameter estimates along the diagonal starting with age-grade combination six-one of Figure 4.12 does however, indicate a certain degree of consistency with this idea. In most

FIGURE 4.16

MALES THREE PARAMETER REGRESSION THROUGH COHORT YEARS

SQUARE OF THE MULTIPLE CORRELATION COEFFICIENT

AGE	GRADE					
	1	2	3	4	5	6
6	0.000000	0.850106	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.978520	0.960561	0.000000	0.000000	0.000000
8	0.000000	0.811612	0.992131	0.760514	0.000000	0.000000
9	0.000000	0.911324	0.338568	0.991706	0.951776	0.000000
10	0.000000	0.729570	0.836769	0.028511	0.991981	0.914964
11	0.000000	0.538546	0.931780	0.970771	0.739532	0.996938
12	0.000000	0.752292	0.889384	0.976998	0.945266	0.946763
13	0.000000	0.840282	0.872752	0.923630	0.976030	0.983826
14	0.000000	0.473337	0.849845	0.760372	0.980194	0.986897
15	0.000000	0.000000	0.896246	0.775381	0.809401	0.960695
16	0.000000	0.000000	0.000000	0.736642	0.097364	0.712397
17	0.000000	0.000000	0.000000	0.000000	0.000000	0.113223
18	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
20	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

	7	8	9	10	11	12
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	0.926872	0.000000	0.000000	0.000000	0.000000	0.000000
12	0.996465	0.930253	0.000000	0.000000	0.000000	0.000000
13	0.965208	0.997102	0.962975	0.000000	0.000000	0.000000
14	0.965098	0.946262	0.997463	0.952418	0.000000	0.000000
15	0.948553	0.946203	0.966007	0.994378	0.957205	0.000000
16	0.798220	0.369871	0.492480	0.979318	0.994169	0.972599
17	0.764835	0.783073	0.854652	0.727075	0.972352	0.993038
18	0.237984	0.016926	0.815308	0.840593	0.885996	0.982745
19	0.000000	0.000000	0.399406	0.435260	0.414076	0.981059
20	0.000000	0.000000	0.000000	0.061551	0.611576	0.933852
21	0.000000	0.000000	0.000000	0.000000	0.609699	0.949682

cases the estimates of β_1 do not permit estimates of grade one totals as each estimate is obtained as a function of the enrollment total for the previous grade.

4.3.2 Two-Cell Averages

The two-cell average is a simple method of estimating enrollment totals by taking the average of two surrounding cells. This method is applied in two ways: (i) averaging along cohort diagonals, and (ii) averaging between cohort years.

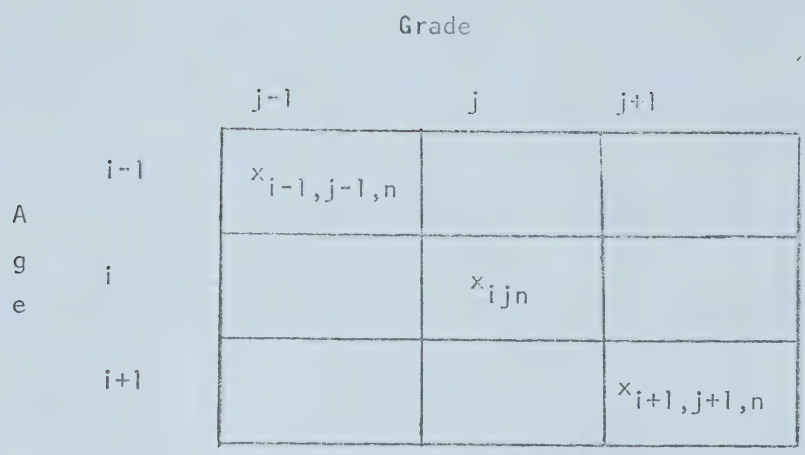
Two-Cell Averages Along Cohort Diagonals

If x_{ijn} is the enrollment total to be estimated, its estimate is obtained from the two-cell average along the cohort diagonal by:

$$\hat{x}_{ijn} = \frac{x_{i-1,j-1,n} + x_{i+1,j+1,n}}{2} , \qquad (4.12)$$

where $x_{i-1,j-1,n}$ and $x_{i+1,j+1,n}$ are the cell totals corresponding to the previous age and grade and the succeeding age and grade for x_{ijn}, and $\hat{x}_{ijn}$ is its estimate. The relationship between the cells is illustrated by Figure 4.17.

FIGURE 4.17



Two-Cell Average Between Cohort Years

This method estimates each enrollment total as the simple average of the enrollment totals for the same age-grade combination for the previous and following cohort year. The estimates are given by

$$x_{ijn} = \frac{x_{i,j,n-1} + x_{i,j,n+1}}{2} . \qquad (4.13)$$

The results of the two-cell average estimation procedures are summarized by tabulating the estimates in cohort table form with the differences between the actual and estimated cell totals. The error of estimate is mainly a test of linearity for the data.

Figure 4.18 is an example of the output for the two-cell average along cohort diagonals for male cohorts 18 in year 1958-59.

FIGURE 4.18

ESTIMATE TWO CELL AVERAGE ALONG COHORT DIAGONAL
FOR MALE COHORTS 18 IN YEAR 58-59

AGE	GRADE										
	1	2	3	4	5	6	7	8	9	10	11
6	0	1364	41	0	0	0	0	0	0	0	0
	0	181	-34	0	0	0	0	0	0	0	0
7	0	4572	1831	0	0	0	0	0	0	0	0
	0	-874	-75	0	0	0	0	0	0	0	0
8	0	2429	3092	1855	154	0	0	0	0	0	0
	0	-17	43	263	-154	0	0	0	0	0	0
9	0	476	2485	2827	2134	0	0	0	0	0	0
	0	-58	243	-341	-180	0	0	0	0	0	0
10	0	154	633	2500	2514	1993	125	0	0	0	0
	0	-18	62	58	6	158	-48	0	0	0	0
11	0	79	197	954	2225	2547	1916	88	0	0	0
	0	-17	23	-105	47	-4	117	-65	0	0	0
12	0	46	83	244	1005	1946	2610	1626	0	0	0
	0	-14	24	15	208	-54	-35	55	0	0	0
13	0	22	36	101	284	1341	1902	2625	1549	51	0
	0	0	33	4	-15	-179	-281	53	-330	-51	0
14	0	9	22	60	94	348	1196	1551	2419	1070	0
	0	1	9	-19	1	-39	273	362	256	348	0
15	0	0	8	19	31	99	326	1339	1444	2373	1012
	0	0	8	3	20	-16	101	-109	37	-213	-91
16	0	0	0	0	12	39	69	408	1044	1313	2286
	0	0	0	0	-4	-17	35	-64	166	-337	-214
17	0	0	0	0	0	7	15	88	301	850	1267
	0	0	0	0	0	-5	13	-33	88	8	-122
18	0	0	0	0	0	0	0	19	48	257	653
	0	0	0	0	0	0	0	-11	25	1	-162
19	0	0	0	0	0	0	0	3	8	48	201
	0	0	0	0	0	0	0	-3	2	-7	-76
20	0	0	0	0	0	0	0	0	0	12	57
	0	0	0	0	0	0	0	0	0	-3	-33

The rows for each age give the estimated enrollment totals and the error of estimate for grades 2 to 11. This method does not permit estimates to be made for grades 1 and 12.

4.3.3 Auto-Regressive Model

Initially auto-regressive models that included terms up to lag five were considered and fitted. However significance tests of the lagged terms revealed that terms of lag greater than one were generally insignificant, therefore, the description of the model will concentrate on auto-regression of lag one.

The general form of the lag one auto-regressive model is

$$x_t = \beta_0 + \beta_1 x_{t-1} + e_t \quad ,$$

where x_t and x_{t-1} are observations at time t and time $t-1$, and e_t is the model error. This model was applied to the school enrollment data in two ways: auto-regression through cohort year, and auto-regression along cohort diagonals. Estimates of the unknown parameters were obtained by the method of least squares, although one of the assumptions of the method is independence of error terms. This assumption does not hold for the auto-regressive model as x_t is related to x_{t-1} which is dependent on e_{t-1}, and hence so is x_t. It can be shown, however, that the least squares estimates are consistent and hence asympototically unbiased. That is, as the sample gets larger the estimate bias vanishes.

4.3.3.1 Auto-Regression Through Cohort Year

The model in this case is

$$x_{ijn} = \beta_0 + \beta_1 x_{i,j,n-1} + e_{ijn} . \qquad (4.14)$$

Each age-grade enrollment total for a given cohort year is expressed as a function of the enrollment total for the same age-grade combination and previous cohort year.

4.3.3.2 Auto-Regression Along Cohort Diagonals

Here, the model is

$$x_{ijn} = \beta_0 + \beta_1 x_{i-1,j-1,n} + e_{ijn} . \qquad (4.15)$$

That is, each enrollment total for age i and grade j is expressed in terms of the age-grade total for the previous age $i-1$ and previous grade $j-1$, and the same cohort year n.

The first model provides a measure of the relationship between age-grade totals from year to year and the second model provides a measure of the relationship between successive age-grade combinations. The results of the fittings are tabled and illustrated, in the same manner for the two- and three-parameter regression, with estimates of the parameters β_0 and β_1, their standard errors and t values along with the actual and estimated enrollment totals and plots.

4.3.4 Giant Regression

The giant regression is an extension of the polynomial regression on age, grade, and year with the inclusion of enrollment totals for the previous age and grade and the previous age and the same grade as predictors for each enrollment total. Thus the giant regression is auto-regressive in nature.

The general form of the model is

$$x_{ijn} = \alpha_0 + \alpha_1 i + \alpha_2 j + \alpha_3 n + \alpha_4 ij + \ldots + \qquad (4.16)$$
$$+ (\beta_0+\beta_1 i+\beta_2 j+\beta_3 n+\beta_4 ij+\ldots)\, x_{i-1,j-1,n} + \ldots +$$
$$+ (\gamma_0+\gamma_1 i+\gamma_2 j+\gamma_3 n+\gamma_4 ij+\ldots)\, x_{i-1,j,n} + e_{ijn} \quad .$$

Stepwise regression was used to select the best predictors from a set that contained terms up to the third order in age, grade, and year. This produced a single equation for estimating the enrollment totals by age, grade, and year for each sex. The fittings were summarized in the same manner as that for the polynomial regression with the predictors tabled in their order of inclusion in the equation.

A multiple correlation coefficient of 0.99 was obtained for both equations with the variables age, grade and year and their squares being the first variables entered as predictors into the equation by the stepwise procedure. The high correlations obtained for this model are due to the auto-regressive terms.

4.4 AGE AND GRADE DISTRIBUTIONS

From the yearly age-grade enrollment tables, tables of the yearly age distribution by grade and the yearly grade distributions by age were constructed for male and female students for the school years 1945-46 to 1969-70. Accompanying each table is a table of the percentage enrollment by grade or age, as the case may be, for each of the school years.

Figures 4.19 to 4.22 are examples of these tables. Figure 4.19 gives the yearly grade distribution for 12 year old males and Figure 4.20 gives the yearly percentages for this age group. Figure 4.20 shows the marked increase in the percentage of 12 year olds in grade 7, from as low as 18 per cent in 1945-46 to almost 60 per cent in 1969-70. This increase is an indication of the accelerated progress of students through the school system over the past 25 years. Figures 4.21 and 4.22 give us another way of viewing the data by tabulating the yearly age distribution for each grade. The particular case shown here is the yearly age distribution of grade 9 females with yearly age percentages. In this case we can see that the average age of grade 9 females has tended to decrease over the years. In 1945-46, 30.4 per cent of the students were 14 years old and 43.6 per cent were 15 years or older, whereas by 1969-70, 58.5 per cent of the students were 14 years old and only 14.4 per cent were 15 years or older, which shows a marked increase in the percentage of students obtaining grade 9 by their fourteenth year.

FIGURE 4.19

YEARLY GRADE DISTRIBUTION FOR 12 YEAR OLD MALES

YEAR	GRADE													TOTAL
	-K-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	
45-46	0	14	34	142	468	1740	1662	1259	1564	57	0	0	0	6940
46-47	0	14	34	132	424	1573	1557	1384	1528	91	0	0	0	6737
47-48	0	14	33	124	402	1625	1697	1377	1613	57	0	0	0	6942
48-49	0	14	32	121	343	1589	1705	1411	1579	57	0	0	0	6851
49-50	0	14	32	116	328	1518	1615	1536	1609	88	0	0	0	6856
50-51	0	14	32	112	294	1349	1643	1989	1614	57	0	0	0	7104
51-52	0	14	31	108	271	1262	1791	2170	1690	117	0	0	0	7454
52-53	0	14	32	107	259	1213	1802	2575	1681	100	0	0	0	7873
53-54	0	14	32	105	248	1134	1907	3172	1690	56	0	0	0	8358
54-55	0	14	32	105	224	1108	2046	3483	1645	56	0	0	0	8713
55-56	0	14	32	103	222	1025	1981	4207	1807	56	0	0	0	9447
56-57	0	14	32	103	226	997	1907	4444	1766	56	0	0	0	9545
57-58	0	14	32	102	218	1021	2015	4494	1919	56	0	0	0	9871
58-59	0	14	32	102	213	995	2304	5252	1876	56	0	0	0	10840
59-60	0	14	32	102	220	996	2463	6285	2174	117	0	0	0	12403
60-61	0	14	33	102	216	952	2177	6477	2388	56	0	0	0	12415
61-62	0	11	27	84	168	890	2429	6924	2178	155	0	0	0	12866
62-63	0	11	30	77	173	830	2376	7065	2740	54	0	0	0	13356
63-64	0	11	28	84	155	722	2210	7400	2836	73	0	0	0	13519
64-65	0	11	20	72	167	806	2458	8236	2671	72	0	0	0	14513
65-66	0	7	24	50	164	667	2476	8747	2932	90	0	0	0	15157
66-67	0	6	30	78	152	732	2620	9306	3094	123	0	0	0	16141
67-68	0	9	21	69	154	669	2629	9726	3372	118	0	0	0	16767
68-69	0	0	6	21	88	634	2627	9876	3667	121	0	0	0	17040
69-70	0	0	6	10	74	547	2742	10468	3536	150	7	0	0	17540

FIGURE 4.20

YEARLY GRADE PERCENTAGES FOR 12 YEAR OLD MALES

YEAR	GRADE												
	-K-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-
45-46	0.0	0.2	0.5	2.0	6.7	25.1	23.9	18.1	22.5	0.8	0.0	0.0	0.0
46-47	0.0	0.2	0.5	2.0	6.3	23.3	23.1	20.5	22.7	1.4	0.0	0.0	0.0
47-48	0.0	0.2	0.5	1.8	5.8	23.4	24.4	19.8	23.2	0.8	0.0	0.0	0.0
48-49	0.0	0.2	0.5	1.8	5.0	23.2	24.9	20.6	23.0	0.8	0.0	0.0	0.0
49-50	0.0	0.2	0.5	1.7	4.8	22.1	23.6	22.4	23.5	1.3	0.0	0.0	0.0
50-51	0.0	0.2	0.5	1.6	4.1	19.0	23.1	28.0	22.7	0.8	0.0	0.0	0.0
51-52	0.0	0.2	0.4	1.4	3.6	16.9	24.0	29.1	22.7	1.6	0.0	0.0	0.0
52-53	0.0	0.2	0.4	1.4	3.3	15.4	24.0	32.7	21.4	1.3	0.0	0.0	0.0
53-54	0.0	0.2	0.4	1.3	3.0	13.6	22.8	38.0	20.2	0.7	0.0	0.0	0.0
54-55	0.0	0.2	0.4	1.2	2.6	12.7	23.5	40.0	18.9	0.6	0.0	0.0	0.0
55-56	0.0	0.1	0.3	1.1	2.3	10.9	21.0	44.5	19.1	0.6	0.0	0.0	0.0
56-57	0.0	0.1	0.3	1.1	2.4	10.4	20.0	46.6	18.5	0.6	0.0	0.0	0.0
57-58	0.0	0.1	0.3	1.0	2.2	10.3	20.4	45.5	19.4	0.6	0.0	0.0	0.0
58-59	0.0	0.1	0.3	0.9	2.0	9.2	21.2	48.4	17.3	0.5	0.0	0.0	0.0
59-60	0.0	0.1	0.3	0.8	1.8	8.0	19.9	50.7	17.5	0.9	0.0	0.0	0.0
60-61	0.0	0.1	0.3	0.8	1.7	7.7	17.5	52.2	19.2	0.5	0.0	0.0	0.0
61-62	0.0	0.1	0.2	0.7	1.3	6.9	18.9	53.8	16.9	1.2	0.0	0.0	0.0
62-63	0.0	0.1	0.2	0.6	1.3	6.2	17.8	52.9	20.5	0.4	0.0	0.0	0.0
63-64	0.0	0.1	0.2	0.6	1.1	5.3	16.3	54.7	21.0	0.5	0.0	0.0	0.0
64-65	0.0	0.1	0.1	0.5	1.2	5.6	16.9	56.7	18.4	0.5	0.0	0.0	0.0
65-66	0.0	0.0	0.2	0.3	1.1	4.4	16.3	57.7	19.3	0.6	0.0	0.0	0.0
66-67	0.0	0.0	0.2	0.5	0.9	4.5	16.2	57.7	19.2	0.8	0.0	0.0	0.0
67-68	0.0	0.1	0.1	0.4	0.9	4.0	15.7	58.0	20.1	0.7	0.0	0.0	0.0
68-69	0.0	0.0	0.0	0.1	0.5	3.7	15.4	58.0	21.5	0.7	0.0	0.0	0.0
69-70	0.0	0.0	0.0	0.1	0.4	3.1	15.6	59.7	20.2	0.9	0.0	0.0	0.0

FIGURE 4.21

YEARLY AGE DISTRIBUTION FOR GRADE 9 FEMALES

YEAR	AGE																	TOTAL
	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	-13-	-14-	-15-	-16-	-17-	-18-	-19-	-20-	-21-	
45-46	0	0	0	0	0	0	0	135	1406	1801	1449	945	171	24	0	0	0	5931
46-47	0	0	0	0	0	0	0	140	1371	1657	1595	1022	174	24	0	0	0	5983
47-48	0	0	0	0	0	0	0	131	1490	1624	1433	965	181	24	0	0	0	5858
48-49	0	0	0	0	0	0	0	134	1427	1686	1553	961	175	24	0	0	0	5960
49-50	0	0	0	0	0	0	0	149	1541	1887	1459	969	176	23	0	0	0	6204
50-51	0	0	0	0	0	0	0	75	1458	2074	1470	929	169	23	0	0	0	6198
51-52	0	0	0	0	0	0	0	164	1562	2221	1461	910	171	23	0	0	0	6512
52-53	0	0	0	0	0	0	0	131	1552	2459	1429	876	166	22	0	0	0	6635
53-54	0	0	0	0	0	0	0	75	1605	2815	1473	854	157	23	0	0	0	7002
54-55	0	0	0	0	0	0	0	157	1782	3152	1369	821	155	22	0	0	0	7458
55-56	0	0	0	0	0	0	0	75	1770	3403	1522	817	157	22	0	0	0	7766
56-57	0	0	0	0	0	0	0	228	1871	3693	1615	827	155	22	0	0	0	8411
57-58	0	0	0	0	0	0	0	75	1776	4434	1787	849	157	22	0	0	0	9150
58-59	0	0	0	0	0	0	0	250	2114	4156	1712	848	161	22	0	0	0	9263
59-60	0	0	0	0	0	0	0	75	1949	4883	1685	788	152	22	0	0	0	9554
60-61	0	0	0	0	0	0	0	76	2069	5918	1800	800	153	22	0	0	0	10838
61-62	0	0	0	0	0	0	0	0	1976	7462	1989	755	139	20	0	0	0	12341
62-63	0	0	0	0	0	0	12	719	3853	5409	1639	681	129	18	0	0	0	12460
63-64	0	0	0	0	0	0	0	0	2644	6351	1974	789	159	33	0	0	0	12450
64-65	0	0	0	0	0	0	0	143	2864	7177	1859	603	127	24	0	0	0	12797
65-66	0	0	0	0	0	0	0	113	3062	7653	1879	695	123	26	0	0	0	13551
66-67	0	0	0	0	0	0	0	172	3574	8107	1929	595	124	16	0	0	0	14517
67-68	0	0	0	0	0	0	0	185	3729	8442	1780	477	72	25	0	0	0	14710
68-69	0	0	0	0	0	0	0	190	4028	9008	1790	396	79	24	9	0	0	15524
69-70	0	0	0	0	0	0	0	170	4234	9503	1845	393	79	20	0	0	8	16252

FIGURE 4.22

YEARLY AGE PERCENTAGES FOR GRADE 9 FEMALES

YEAR	AGE													
	-5-	-6-	-7-	-8-	-9-	-10-	-11-	-12-	-13-	-14-	-15-	-16-	-17-	-18-
45-46	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3	23.7	30.4	24.4	15.9	2.9	0.4
46-47	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.3	22.9	27.7	26.7	17.1	2.9	0.4
47-48	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	25.4	27.9	24.5	16.5	3.1	0.4
48-49	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	23.9	28.3	26.1	16.1	2.9	0.4
49-50	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.4	24.9	30.4	23.5	15.6	2.8	0.4
50-51	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	23.5	32.5	23.7	15.0	2.7	0.4
51-52	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	24.0	34.1	22.4	14.0	2.6	0.4
52-53	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	23.4	37.1	21.5	13.2	2.5	0.3
53-54	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	22.9	40.2	21.0	12.2	2.2	0.3
54-55	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.1	23.9	42.3	18.4	11.0	2.1	0.3
55-56	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	22.5	44.5	19.4	10.4	2.0	0.3
56-57	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	22.2	43.9	19.2	9.8	1.8	0.3
57-58	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	19.4	49.0	19.5	9.3	1.7	0.2
58-59	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	22.8	44.9	18.5	9.2	1.7	0.2
59-60	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	20.4	51.1	17.6	8.2	1.6	0.2
60-61	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	19.1	54.6	16.6	7.4	1.4	0.2
61-62	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.0	60.5	16.1	6.1	1.1	0.2
62-63	0.0	0.0	0.0	0.0	0.0	0.0	0.1	5.8	30.9	43.4	13.2	5.5	1.0	0.1
63-64	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.2	55.0	15.9	6.3	1.3	0.3
64-65	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	22.4	56.1	14.5	4.7	1.0	0.2
65-66	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	22.6	56.5	13.9	5.1	0.9	0.2
66-67	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	24.6	55.8	13.3	4.1	0.9	0.1
67-68	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	25.4	57.4	12.1	3.2	0.5	0.2
68-69	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	25.9	58.0	11.5	2.6	0.5	0.2
69-70	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	26.1	58.5	11.4	2.4	0.5	0.1

As can be seen, these tables can be very useful in the direction of trends in enrollments that can only be detected with considerable difficulty in the original annual enrollment tables.

CHAPTER V

PROJECTION BY MULTI-CHANNEL TIME SERIES ANALYSIS

5.1 INTRODUCTION

Time series analysis is a technique for categorizing and studying movements in data that consist of successive values of a variable measured at specified time intervals. The time series employed is a stochastic function, student enrollment, acted upon by the independent variable, time. The time interval consists of one year periods, each containing enrollment totals by age-grade-sex for 24 intervals. A discrete time series with t taking on integer values 0 to 24 is used with our function being enrollment in each of the 25 years. The data used are non-experimental, that is, we can only observe the enrollment in a school system. It would be difficult, if not impossible, to conduct planned experiments on student enrollment.

Ideally our time series should measure the same thing in the same way throughout its period of coverage. However, the enrollment data exhibits various inconsistencies, particularly in those years where policy decisions affected enrollment as discussed in an earlier chapter. The refinement of the data (Appendix A), which was undertaken at that point, provides us with an improved data base upon which we employ our model.

We use time series analysis both to study past patterns and to project the long-term sweep of the data, usually referred

to as the secular trends of the series. The past pattern of the series is studied in order to obtain a better picture of the variables that affect our enrollment. This gives us an indication of the future direction the pattern may reasonably be expected to take. Our primary goal is to extrapolate or project the trend curve to estimate its level for some future period on the basis of its pattern of past behaviour.

Our choices of a relevant time period and of the appropriate type of trend curve are greatly affected by the assumptions made regarding the future course of events.

5.2 MODEL

For the purpose of this study we consider two types of multi-channel time series that will be classified as grade specific and age specific series. The grade specific series is a twelve channel series with each channel (variable) representing the number of students in each of the twelve grades during each year. The age specific series is a fourteen channel series with each channel representing the number of students age 5, 6, ..., 17, and 18+ years. If we denote the grade or age specific series by a random vector $\underline{y}$ of length p, where p is the number of channels, i.e., 12 or 14 channels respectively, then the data may be arranged in a time series of the form,

$$\underline{y}_0, \underline{y}_1, \cdots, \underline{y}_t, \cdots, \underline{y}_n \qquad (5.1)$$

where $\underline{y}_t$, enrollment, represents the value of the random vector

$\underline{y}$ at time t measured in years.

The problem is to make a prediction of the vector, i.e., a projection that is a prediction of $\underline{y}_{n+\alpha}$, where $\alpha = 1,2,3,\ldots$, based on the series given by equation (5.1). Generally we would not expect this series to be stationary, that is, in statistical equilibrium over time, due to factors such as overall population trends, change of school system policies, and the tendency for students to stay in school longer. It is also likely that many factors tend to interact with each other in unknown patterns. We know very little about how these factors acting upon the enrollment data cause non-stationarity.

To make the problem manageable we assume that the annual increase in enrollment for each age or grade is, on the average, the same for a long range of years. As a consequence, if we take the successive difference of the given time series, equation (5.1) and create a new series, then this new time series, a difference series, is stationary. Box and Jenkins (1968) have suggested applicability of such techniques for uni-channel time series. Although higher order difference series are also possible, we limit ourselves to the first order time series. Under this restrictive assumption we will apply the time series techniques to the enrollment projection problem.

The first order difference series is defined as,

$$\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_n \quad , \tag{5.2}$$

where $\underline{d}_t = \underline{y}_t - \underline{y}_{t-1}$, for $t = 1,2,\ldots,n$. $\underline{d}_t$ thus expresses

the difference in enrollment between any two successive years. From this difference series we create a deviation series by removing the mean vector of $\{\underline{d}_t\}$, denoted by $\bar{\underline{d}}_.$. We write this deviation series as

$$\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_n \quad , \tag{5.3}$$

where $\underline{x}_t = \underline{d}_t - \bar{\underline{d}}_.$, and $\bar{\underline{d}}_. = (\sum_{t=1}^{n} \underline{d}_t)/n$. The expected values of each term of the new series is equal to a null vector.

The model used for projection is a multivariate version of an auto-regressive scheme which was discussed earlier in the data analysis chapter,

$$\underline{x}_t = \underline{F}_1 \underline{x}_{t-1} + \underline{F}_2 \underline{x}_{t-2} + \cdots + \underline{F}_m \underline{x}_{t-m} + \underline{e}_t \quad , \tag{5.4}$$

where $\{\underline{F}_r\}$ are the filter matrices. Each filter matrix is a p by p constant or a regression coefficient matrix of $\underline{x}_t$ on $\underline{x}_{t-r}$, where $r = 1,2,\ldots,m$. The constant, m, is the order of the scheme and is defined as being the maximum possible time lag on which $\underline{x}_t$ depends, and $\underline{e}_t$ is the random component, fluctuating due to "noise" or "error" found in the system. From this it can be seen that $\underline{x}_t$ depends on the past series only through $\underline{x}_{t-1}, \underline{x}_{t-2}, \cdots, \underline{x}_{t-m}$, or up to the maximum time lag m. The simplest case occurs when $m = 1$ and the model is then known as first order Markovian, i.e.,

$$\underline{x}_t = \underline{F}_1 \underline{x}_{t-1} + \underline{e}_t \quad . \tag{5.5}$$

We are interested in projecting future years, say, α years ahead for $\underline{x}_t$ from our reference point, and therefore consider a time series which is α units of time ahead of $\underline{x}_t$, or,

$$\underline{z}_t = \underline{x}_{t+\alpha} \quad \text{for} \quad t = -\alpha+1, -\alpha+2, \ldots, -1, 0, 1, \ldots, n-\alpha \ .$$

This new time series, $\{\underline{z}_t\}$, is an output series while the series $\{\underline{x}_t\}$ is called an input series.

For $t < -\alpha+1$ or $t > n-r$, we have no information on $\underline{z}_t$ or $\underline{x}_{t+\alpha}$. Thus these values are not defined in these regions of time. We will make the simplifying assumption that in these regions the series takes the mean of the series or that of a null vector since the $\{\underline{x}_t\}$ is created by removing the mean vector from $\{\underline{d}_t\}$. This assumption is justifiable only if we can assume that the input series $\{\underline{x}_t\}$ and the output series $\{\underline{z}_t\}$ are stationary. Under these conditions

$$\underline{z}_t = \underline{x}_{t+\alpha} = \underline{0}, \quad \text{for} \quad \begin{cases} t < -\alpha+1 \\ \\ t > n-\alpha \end{cases} \ . \tag{5.6}$$

We can then rewrite the equation of the model, (5.4), as

$$\underline{z}_t = \underline{x}_{t+\alpha} = \sum_{r=1}^{m} \underline{F}_r \, \underline{x}_{t+\alpha-r} + \underline{e}_t \ , \tag{5.7}$$

where $-\infty < t < \infty$. The estimate of $\underline{F}_r$ may be obtained by the usual least square methods. We find that

$$\hat{\underline{z}}_t = \sum_{r=1}^{m} \hat{\underline{F}}_r \underline{x}_{t+\alpha-r} \quad , \tag{5.8}$$

where $\hat{\underline{z}}_t$ and $\hat{\underline{F}}_r$ are the least square estimates of $\underline{z}_t$ and $\underline{F}_r$, respectively.

5.3 ASSUMPTIONS OF THE MODEL

The basic assumptions, underlying the model are:

(a) Input $\{\underline{x}_t\}$ and output $\{\underline{z}_t\}$ series are stationary. Stationary series contain no trends, they are in statistical equilibrium. This means the statistical properties (such as distributions, moments) of the input and output remain invariant over time.

(b) The estimation of $\{\underline{z}_t\}$ is based on the least square method. We estimate the filters $\underline{F}_r$ in such a way as to minimize the trace of the mean-square-error matrix between the output $\{\underline{z}_t\}$ and its estimate $\{\hat{\underline{z}}_t\}$, i.e.,

$$\text{minimize the trace of } \sum_{t=-\alpha+1}^{n-\alpha} (\underline{z}_t - \hat{\underline{z}}_t)(\underline{z}_t - \hat{\underline{z}}_t)' \quad .$$

(c) The operation to be used for the projection is assumed to be a linear operation on available information consisting of the past history of the input series.

(d) The unknown portion of the series $\{\underline{x}_t\}$ or $\{\underline{z}_t\}$ are equal to the mean of the series, i.e., $\underline{0}$, as given by equation (5.6).

It is obvious that neighbouring values of input and output series will be influenced by each other. Besides stating the mean and variance we have to specify the sample auto-covariance of each series. The cross-covariance function shows the interactions between the two series. Let $\underline{R}_{xx}(r)$, $\underline{R}_{zx}(r)$, and $\underline{R}_{zz}(r)$ be the sample auto-covariance of $\underline{x}$, cross-covariance between $\underline{z}$ and $\underline{x}$, and auto-covariance of $\underline{z}$, respectively, based on sample size n with time lag r, where $r = -m, -m+1, \ldots, 0, 1, \ldots, m$. Then

$$\begin{aligned}
\underline{R}_{xx}(r) &= \text{Cov}\ (\underline{x}_t, \underline{x}_{t-r}) = (\textstyle\sum \underline{x}_t \underline{x}'_{t-r})/n \\
\underline{R}_{zx}(r) &= \text{Cov}\ (\underline{z}_t \underline{x}_{t-r}) = \text{Cov}\ (\underline{x}_{t+\alpha}, \underline{x}_{t-r}) \\
&= (\textstyle\sum \underline{z}_t \underline{x}'_{t-r})/n = (\sum \underline{x}_{t+\alpha} \underline{x}'_{t-r})/n \\
\underline{R}_{zz}(r) &= \text{Cov}\ (\underline{z}_t, \underline{z}_{t-r}) = \text{Cov}\ (\underline{x}_{t+\alpha}, \underline{x}_{t+\alpha-r}) \\
&= (\textstyle\sum \underline{z}_t \underline{z}'_{t-r})/n = (\sum \underline{x}_{t+\alpha} \underline{x}'_{t+\alpha-r})/n \quad ,
\end{aligned} \tag{5.9}$$

where summation is extended to all non-null vectors.

Then it can be shown that the minimum-mean-square method gives a normal equation,

$$\underline{R}\,\hat{\underline{F}} = \underline{G} \tag{5.10}$$

where

$$\underline{R} = \begin{bmatrix} \underline{R}_{xx}(0) & \underline{R}_{xx}(-1) & \underline{R}_{xx}(-2) & \cdots & \underline{R}_{xx}(-m) \\ \underline{R}_{xx}(1) & \underline{R}_{xx}(0) & \cdots & \cdots & \underline{R}_{xx}(1-m) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{R}_{xx}(m) & \underline{R}_{xx}(m-1) & \cdots & \cdots & \underline{R}_{xx}(0) \end{bmatrix}, \tag{5.11}$$

$$\hat{\underline{F}} = \begin{bmatrix} \hat{\underline{F}}_0 \\ \hat{\underline{F}}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\underline{F}}_m \end{bmatrix} \qquad \underline{G} = \begin{bmatrix} \underline{R}_{zx}(0) \\ \underline{R}_{zx}(1) \\ \cdot \\ \cdot \\ \cdot \\ \underline{R}_{zx}(m) \end{bmatrix},$$

and $\hat{\underline{F}}_0$ is a $p \times p$ identity matrix.

The estimates $\hat{F}_1, \hat{F}_2, \ldots, \hat{F}_m$ are given as the solution of equation (5.10),

$$\hat{\underline{F}} = (\underline{R}'\,\underline{R})^{-1}\,\underline{G} \ .$$

5.4 RESULTS

5.4.1 Order of the Scheme

The order of scheme of the series $\{\underline{x}_t\}$, or the maximum possible m, is unknown. However, inferences on the m are possible by using a partial correlation method (Quenuille, 1968) which uses an approximate χ^2 test for large n, the number of years in our time series. For $n = 24$ our series is too short in comparison with the number of channels, $p = 12$ or 14, to use the method directly. For this length of series only the estimation of $\underline{F}_1$ is possible while for a suitable test of $m = 1$ the calculation of $\hat{\underline{F}}_1$, $\hat{\underline{F}}_2$, $\hat{\underline{F}}_3$ are required. We use as an alternative, despite its obvious shortcoming as an inference technique, a test on some subset of channels, or by collapsing the channels such that p is reduced to a manageable number such as 3. This is done in the case of grades by collapsing grades 1 to 4, 5 to 8, and 9 to 12, and for the case of ages by collapsing the ages 5 to 9, 10 to 13, and 14 to 18+.

For the purpose of exploration these tests are performed on the grade and age series. The results are summarized in Figures 5.1 and 5.2. The actual computations are done by the computer program TIME 15 which uses Robinson's (1967) subprograms.

FIGURE 5.1

Test for m = 1, Grade Specific Series

Series	Channel 1	2	3	Chi-Square	Probability*
A	1	2	3	19.443	0.022
B	4	5	6	4.967	0.837
C	7	8	9	13.189	0.154
D	10	11	12	48.521	< 0.001
E	(1~4)	(5~8)	(9~12)	5.833	0.756

* With degrees of freedom $p^2 = 9$.

FIGURE 5.2

Test for m = 1, Age Specific Series

Series	Channel 1	2	3	Chi-Square	Probability*
F	(5+6)	7	8	22.009	0.008
G	9	10	11	29.195	0.001
H	12	13	14	8.687	0.467
I	15	16	(17+18)	8.005	0.534
J	(5~9)	(10~13)	(14~18+)	8.356	0.499

* With degrees of freedom $p^2 = 9$.

We note that the series D (grade specific), involving grades (10, 11, 12), is significant at the $p = 0.01$ level. This is probably the result of more students continuing in these grades in recent years, hence destroying our assumption of the stationarity properties at these particular grade levels.

For the age specific table the series F and G, namely ages 5 to 8 and 9 to 11, are significant at the 0.01 level. This may be the result of the increased numbers of five year olds being admitted to school. The minimum age requirement for entering school has been lowered in recent years.

5.4.2 Filters

If we accept the assumption that the order of the scheme is one (the series is Markovian), then the equation of the model will be,

$$\underline{z}_t = \underline{x}_{t+1} = \underline{F}_1 \underline{x}_t + \underline{e}_t \quad ,$$

or (5.12)

$$\underline{d}_{t+1} = \underline{F}_1 [\underline{d}_t - \bar{\underline{d}}] + \bar{\underline{d}} + \underline{e}_t \quad .$$

with $\alpha = 1$, i.e., a projection of enrollment one year ahead.

The estimation of $\underline{F}_1$ from a given series $\{\underline{y}_t\}$ of length $(n+1)$ is obtained by a computer program named TIME 10 which uses Robinson's (1967) multi-channel time series subroutines together with a number of input/output and other subprograms. The

equivalent computation may be done by any other multivariate regression program using the least square method (e.g., MULV55, DERS, 1969).

With an estimated filter matrix $\hat{\underline{F}}_1$ and a mean vector $\bar{\underline{d}}$ from the computer program, we generate an estimate of the output series, given an initial value $\underline{d}_0$,

$$\hat{\underline{d}}_t = \hat{\underline{F}}_1 (\hat{\underline{d}}_{t-1} - \bar{\underline{d}}) + \bar{\underline{d}} \quad , \tag{5.13}$$

in successive series. These, in turn, can be used to generate the estimate of the original output series $\{\underline{y}_t\}$, namely,

$$\hat{\underline{y}}_t = \hat{\underline{y}}_{t-1} + \hat{\underline{d}}_t \quad , \qquad t = n+1, n+2, \ldots \quad .$$

The above operation is programmed in a computer program called HRRC 02. The filter matrices for the grade and age specific series, along with the mean vector $\bar{\underline{D}}$, are given in Figures 5.3 and 5.4

FIGURE 5.3

Filter and Mean Estimates, Grade Specific, 1946-70

Age \ Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18+
5	-0.753	0.763	-1.049	0.389	-0.308	0.010	0.227	0.447	0.098	0.115	-0.150	-0.073	0.141	-1.586
6	0.716	0.440	0.362	-0.253	-0.050	0.190	-0.024	0.270	-0.045	0.035	-0.126	-0.187	0.295	-0.784
7	0.780	0.542	0.350	-0.307	-0.043	0.142	-0.142	0.157	0.136	0.159	-0.334	0.088	-0.201	0.030
8	0.059	-0.338	1.125	-0.053	-0.028	-0.027	-0.074	0.214	-0.189	0.133	0.184	-0.078	-0.492	0.239
9	-0.089	-0.558	0.438	0.898	0.134	-0.186	0.062	-0.032	-0.109	0.023	0.281	-0.191	0.099	0.233
10	-0.121	-0.080	0.014	-0.034	0.890	-0.085	0.030	0.026	0.008	0.073	-0.028	-0.142	-0.103	0.148
11	-0.164	0.286	-0.379	0.023	0.009	0.958	0.076	0.120	-0.030	0.103	-0.187	-0.176	0.200	-0.138
12	-0.472	-0.128	0.052	-0.054	-0.013	-0.112	-1.019	-0.205	0.148	0.031	-0.209	0.208	0.144	-0.065
13	0.048	-0.302	0.254	-0.004	-0.014	-0.047	0.007	0.907	0.018	0.128	-0.140	0.176	-0.080	0.114
14	0.112	0.035	-0.265	0.090	-0.206	0.048	0.056	-0.102	1.407	-0.224	-0.434	0.723	-0.315	-0.477
15	-0.026	-0.127	0.014	0.099	0.009	-0.043	0.150	0.090	0.088	0.735	0.141	0.061	-0.215	-0.280
16	0.054	0.159	-0.092	0.069	0.073	-0.026	0.338	-0.011	0.163	0.003	0.642	0.191	-0.427	-0.200
17	0.051	0.226	0.257	0.040	-0.063	0.031	0.184	0.102	0.317	0.003	-0.094	0.811	-0.054	-0.652
18+	-0.045	0.018	0.009	-0.107	0.060	0.060	0.163	0.134	0.111	-0.132	-0.075	-0.085	0.418	-0.185
MEAN	0.390	0.882	0.938	0.939	0.954	0.885	0.851	0.850	0.878	0.793	0.881	0.794	0.517	0.209

FIGURE 5.4

Filter and Mean Estimates, Grade Specific, 1946-70

Grade \ Grade	1	2	3	4	5	6	7	8	9	10	11	12
1	-0.020	-0.344	0.073	0.243	0.020	0.750	-0.180	0.442	-0.406	-0.026	-0.521	0.042
2	0.848	0.153	-0.211	0.048	-0.137	0.094	0.090	-0.093	-0.223	-0.183	0.131	-0.181
3	-0.155	0.759	-0.045	-0.222	0.114	-0.226	0.327	-0.229	0.232	-0.019	0.081	-0.078
4	-0.041	-0.032	0.911	-0.009	-0.025	-0.061	0.129	-0.203	0.119	0.105	-0.155	0.083
5	0.135	-0.072	-0.027	0.781	0.069	-0.205	0.013	-0.055	0.032	0.139	0.055	-0.035
6	-0.042	-0.003	-0.168	0.066	0.080	0.053	0.070	0.047	0.032	0.122	0.091	-0.194
7	-0.006	0.016	-0.044	0.048	0.046	0.994	0.193	-0.099	0.122	-0.047	-0.112	-0.096
8	0.012	-0.016	-0.162	0.167	-0.223	0.208	0.837	0.099	0.111	0.003	0.098	-0.193
9	0.057	0.040	0.027	0.002	0.116	-0.091	-0.033	0.780	-0.011	-0.014	-0.031	0.201
10	-0.046	-0.111	-0.231	0.064	-0.290	0.281	-0.187	0.416	0.936	0.037	0.235	-0.319
11	0.087	0.038	-0.027	0.080	0.019	0.115	-0.068	0.249	-0.044	0.811	0.217	0.336
12	0.112	0.025	0.057	0.020	0.117	-0.021	0.119	0.152	-0.319	0.208	0.787	0.159
MEAN	0.875	0.915	0.907	0.903	0.912	0.897	0.882	0.876	0.915	0.947	0.851	0.883

5.4.3 Projections of Enrollment

Starting with 1971, projections of enrollment were made for every year up to and including the school year 2004-05 for both the age and grade specific series. Results for male and female students are given in Figures 5.5 to 5.8.

Figure 5.5 gives the projections by grade specific time series for males. For example, it is projected that there will be 25,700 male students in grade 5 for the school year 1983-84. The total projected male enrollment as given by the last column is 290,300 for 1983-84. Figure 5.6 gives similar results for females. Total projected enrollment for any year is less for the females than the males because the most recent given enrollments demonstrate a similar situation. Because stationarity principles have been assumed, as well as identical successive differences, projected enrollment increases in a linear fashion. The difference between any two successive years for each grade is either 400 or 500, depending on round off error. This makes each of the figures orderly in their appearance. Enrollment drops steadily for each year across grades until grade 12 where it increases slightly. This is explained by the inclusion of former grade 12 students who have returned for additional course credits.

Figures 5.7 and 5.8 give the projections by age specific time series for males and females. For example it is projected that there will be 25,500 male students age 9 in 1983-84. This corresponds very closely to 25,700 male students in grade 5 to be the "proper" grade for 9 year olds we can see that the age and

FIGURE 5.5

Projection by Grade Specific Time Series, Male

(Unit = 100)

Year \ Grade	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL
1970 - 1971	205	200	202	199	197	191	189	185	167	157	148	151	2191
1971 - 1972	210	203	204	204	201	195	192	189	174	162	152	160	2247
1972 - 1973	214	208	208	206	206	199	196	194	180	170	155	162	2299
1973 - 1974	216	212	214	210	210	205	200	198	184	175	161	166	2351
1974 - 1975	222	215	218	216	214	208	205	202	188	181	166	171	2407
1975 - 1976	226	220	221	220	220	213	208	207	193	185	171	177	2462
1976 - 1977	230	225	226	223	225	219	213	210	199	190	175	182	2516
1977 - 1978	234	229	231	228	228	223	218	215	202	195	179	186	2569
1978 - 1979	239	234	235	233	233	227	223	220	206	200	183	190	2623
1979 - 1980	244	239	240	237	238	232	226	224	212	204	188	195	2679
1980 - 1981	248	244	245	242	243	237	231	228	217	210	192	199	2734
1981 - 1982	252	249	250	247	247	241	236	233	220	215	196	204	2790
1982 - 1983	257	253	254	251	252	246	240	238	226	219	201	209	2847
1983 - 1984	262	258	259	256	257	251	245	242	231	224	205	214	2903
1984 - 1985	266	263	264	260	262	255	250	247	235	230	209	218	2959
1985 - 1986	270	267	269	265	267	260	254	252	240	234	214	223	3015
1986 - 1987	275	272	273	270	272	265	259	256	245	240	218	228	3072
1987 - 1988	280	277	278	274	277	270	264	261	250	245	223	232	3129
1988 - 1989	284	281	283	279	281	274	269	266	255	250	227	237	3185
1989 - 1990	288	286	287	284	286	279	273	270	259	255	232	242	3242
1990 - 1991	293	291	292	288	291	284	278	275	264	260	236	246	3299
1991 - 1992	297	295	297	293	296	289	282	280	269	265	241	251	3355
1992 - 1993	302	300	301	298	301	294	287	284	274	270	245	256	3412
1993 - 1994	306	305	306	302	305	298	292	289	279	275	250	261	3469
1994 - 1995	311	309	311	307	310	303	297	294	284	280	254	266	3525
1995 - 1996	315	314	315	312	315	308	301	298	289	286	259	271	3582
1996 - 1997	320	319	320	316	320	313	306	303	293	291	263	275	3639
1997 - 1998	324	323	325	321	325	317	311	307	298	296	268	280	3695
1998 - 1999	329	328	329	325	330	322	315	312	303	301	272	285	3752
1999 - 2000	333	333	334	330	334	327	320	317	308	306	277	290	3808
2000 - 2001	337	337	339	335	339	332	325	321	313	311	281	294	3865
2001 - 2002	342	342	343	339	344	336	329	326	318	316	286	299	3922
2002 - 2003	346	347	348	344	349	341	334	331	322	321	290	304	3978
2003 - 2004	351	352	353	349	354	346	339	335	327	326	295	309	4035
2004 - 2005	355	356	358	353	358	351	343	340	332	332	299	314	4092

FIGURE 5.6

Projection by Grade Specific Time Series, Female

(Unit = 100)

Year \ Grade	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL
1970 - 1971	199	190	186	185	181	179	172	171	164	153	141	129	2047
1971 - 1972	201	195	191	187	184	182	175	175	167	158	143	135	2094
1972 - 1973	204	197	196	192	186	185	178	178	172	162	147	139	2138
1973 - 1974	209	200	199	198	192	187	182	181	175	166	150	143	2182
1974 - 1975	214	205	202	201	197	193	184	184	179	170	154	146	2228
1975 - 1976	218	210	207	204	200	198	189	186	182	173	157	149	2274
1976 - 1977	223	214	212	209	204	201	194	191	185	177	160	152	2322
1977 - 1978	228	219	216	213	208	205	197	196	190	179	164	156	2371
1978 - 1979	231	224	220	218	212	208	200	199	195	184	166	159	2419
1979 - 1980	236	229	226	222	217	213	205	203	299	189	171	161	2470
1980 - 1981	240	233	230	227	221	218	209	207	202	193	176	166	2522
1981 - 1982	245	238	234	231	226	222	213	212	206	197	179	171	2573
1982 - 1983	249	242	238	235	230	226	218	216	211	201	183	174	2623
1983 - 1984	253	247	243	240	234	230	222	220	216	206	186	178	2674
1984 - 1985	258	252	248	245	238	234	226	224	220	210	191	181	2726
1985 - 1986	262	256	252	249	243	239	230	228	224	214	195	186	2778
1986 - 1987	266	260	256	253	247	243	234	232	228	219	199	190	2829
1987 - 1988	271	265	260	258	251	247	238	236	233	223	203	194	2880
1988 - 1989	275	269	265	262	256	251	243	241	237	228	207	198	2931
1989 - 1990	279	274	269	266	260	256	247	245	241	232	211	202	2982
1990 - 1991	283	278	274	271	264	260	251	249	246	237	215	206	3034
1991 - 1992	288	282	278	275	268	264	255	253	250	241	219	210	3085
1992 - 1993	292	287	282	280	273	268	259	256	254	245	223	214	3136
1993 - 1994	296	292	287	284	277	272	263	261	259	250	227	219	3187
1994 - 1995	300	296	291	288	281	277	268	265	263	254	231	223	3238
1995 - 1996	305	301	296	293	286	281	272	269	267	258	235	227	3289
1996 - 1997	309	305	300	297	290	285	276	273	272	263	239	231	3339
1997 - 1998	313	309	304	302	294	289	280	276	276	267	243	234	3390
1998 - 1999	318	314	309	306	299	293	284	282	280	272	248	239	3441
1999 - 2000	322	318	313	310	303	297	288	286	284	276	252	243	3493
2000 - 2001	326	323	317	315	307	302	292	290	289	280	255	247	3543
2001 - 2002	331	327	322	319	311	306	296	294	293	285	259	251	3594
2002 - 2003	335	332	326	324	316	310	300	298	297	289	264	255	3645
2003 - 2004	339	336	331	328	320	314	305	302	302	293	268	259	3696
2004 - 2005	343	341	335	332	324	318	309	306	306	298	272	263	3747

FIGURE 5.7

Projection by Age Specific Time Series, Male

(Unit = 100)

Year \ Grade	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970 - 1971	55	191	191	196	192	183	186	182	178	170	157	144	105	66	2195
1971 - 1972	57	196	196	200	196	188	188	187	185	175	163	146	107	68	2252
1972 - 1973	57	200	201	205	200	192	193	189	192	182	167	150	108	70	2308
1973 - 1974	58	203	206	211	205	196	197	195	194	188	175	155	111	69	2364
1974 - 1975	62	206	209	217	212	201	200	198	198	189	182	162	114	70	2421
1975 - 1976	64	211	213	219	218	207	204	201	203	194	184	168	119	72	2478
1976 - 1977	66	216	219	223	219	212	210	206	207	199	190	170	122	74	2532
1977 - 1978	67	221	224	228	224	215	215	211	211	201	194	174	123	76	2585
1978 - 1979	69	225	229	234	229	219	218	217	217	206	198	179	126	77	2641
1979 - 1980	71	230	234	240	235	224	221	219	222	211	204	182	129	78	2699
1980 - 1981	74	234	238	244	240	229	226	223	225	215	209	187	131	79	2755
1981 - 1982	75	238	244	249	245	234	231	228	229	218	214	192	134	80	2810
1982 - 1983	77	243	248	254	250	238	236	233	234	221	218	197	137	82	2867
1983 - 1984	80	247	253	258	255	243	240	237	239	226	222	200	140	84	2923
1984 - 1985	81	253	258	263	259	248	245	241	244	231	226	203	142	86	2979
1985 - 1986	83	257	263	269	264	251	249	246	248	235	232	208	144	87	3035
1986 - 1987	85	261	267	274	270	256	253	250	252	238	237	213	147	88	3093
1987 - 1988	87	265	272	278	275	262	258	254	257	243	241	218	151	90	3150
1988 - 1989	89	270	277	283	279	267	263	259	261	247	246	222	153	91	3206
1989 - 1990	91	275	282	288	284	271	268	264	266	250	251	226	155	93	3262
1990 - 1991	93	279	286	293	289	275	272	268	271	254	255	230	158	94	3319
1991 - 1992	95	284	292	298	294	280	276	272	276	259	260	234	161	96	3376
1992 - 1993	97	288	297	303	299	285	280	276	280	263	265	239	163	97	3432
1993 - 1994	98	293	302	308	304	290	285	281	284	267	270	244	166	98	3489
1994 - 1995	101	297	306	313	309	294	289	286	289	271	275	248	169	100	3546
1995 - 1996	103	302	311	317	314	299	294	290	293	275	279	252	172	102	3602
1996 - 1997	104	306	316	322	318	304	298	294	298	279	284	256	175	103	3659
1997 - 1998	106	311	321	327	323	308	303	299	302	283	289	261	177	104	3715
1998 - 1999	108	315	326	333	329	312	307	303	307	287	294	265	180	106	3772
1999 - 2000	110	319	330	337	334	317	311	308	312	292	299	270	183	107	3829
2000 - 2001	112	324	335	341	338	322	316	312	316	296	303	274	186	109	3885
2001 - 2002	114	329	340	347	343	327	321	317	320	299	308	278	188	110	3942
2002 - 2003	116	333	345	352	348	331	325	321	325	303	313	283	191	112	3999
2003 - 2004	118	338	350	357	354	336	329	325	330	308	318	287	194	113	4055
2004 - 2005	120	342	355	362	358	341	334	330	334	312	323	291	196	115	4112

FIGURE 5.8

Projection by Age Specific Series, Female

(Unit = 100)

Year \ Grade	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970 - 1971	58	184	182	186	187	172	170	173	166	162	151	134	95	29	2049
1971 - 1972	59	189	186	189	190	176	174	177	173	163	154	136	97	30	2094
1972 - 1973	60	193	191	194	193	179	178	180	176	169	158	139	99	30	2138
1973 - 1974	63	196	195	199	198	183	181	184	180	171	162	142	101	31	2184
1974 - 1975	65	201	199	203	204	187	184	187	184	175	165	145	102	31	2233
1975 - 1976	67	205	204	207	208	192	187	191	187	180	169	148	105	31	2281
1976 - 1977	69	210	209	212	212	196	193	194	191	181	173	151	107	32	2330
1977 - 1978	71	214	213	217	216	200	196	199	194	184	176	156	109	33	2379
1978 - 1979	73	219	218	221	221	204	200	203	199	189	180	159	112	33	2430
1979 - 1980	75	223	222	225	226	208	204	207	203	193	184	162	114	34	2480
1980 - 1981	77	228	227	230	230	213	209	211	207	195	188	165	116	34	2530
1981 - 1982	79	232	232	235	235	216	212	215	211	200	192	169	118	35	2581
1982 - 1983	81	236	236	239	240	221	216	219	216	204	196	173	121	36	2633
1983 - 1984	83	241	241	243	244	225	221	223	220	207	200	176	123	36	2684
1984 - 1985	84	245	246	248	248	229	225	227	224	211	204	180	126	37	2734
1985 - 1986	87	249	250	253	253	233	229	232	228	215	208	183	128	37	2786
1986 - 1987	89	254	254	257	258	238	232	236	233	219	212	187	131	38	2837
1987 - 1988	90	258	259	262	262	242	237	239	237	223	216	190	133	39	2888
1988 - 1989	92	262	264	266	267	246	241	244	241	227	220	194	136	39	2939
1989 - 1990	95	267	268	271	272	250	245	248	245	230	224	198	138	40	2990
1990 - 1991	96	271	273	275	276	255	249	252	250	235	228	201	141	41	3041
1991 - 1992	98	275	277	280	281	259	253	256	254	238	232	205	143	41	3092
1992 - 1993	100	280	282	284	285	263	257	260	258	242	236	209	145	42	3143
1993 - 1994	102	284	286	289	290	267	261	264	262	246	240	212	148	43	3194
1994 - 1995	104	288	291	293	295	271	265	268	267	250	244	216	150	43	3245
1995 - 1996	106	293	295	298	299	275	269	272	271	254	248	219	153	44	3296
1996 - 1997	108	297	300	302	304	280	273	276	275	257	252	223	155	44	3347
1997 - 1998	110	301	304	307	309	284	277	280	279	261	256	226	158	45	3398
1998 - 1999	112	306	309	311	313	288	281	285	283	265	260	230	160	46	3449
1999 - 2000	114	310	314	316	318	292	285	289	288	269	264	234	162	46	3500
2000 - 2001	116	314	318	320	322	296	289	293	292	273	268	237	165	47	3551
2001 - 2002	118	319	323	325	327	301	293	297	296	277	272	241	167	48	3602
2002 - 2003	120	323	327	329	332	305	297	301	300	281	276	244	170	48	3653
2003 - 2004	122	327	332	334	336	309	301	305	304	284	280	248	172	49	3704
2004 - 2005	124	332	336	338	341	313	305	309	309	288	284	252	175	50	3755

grade specific tables correspond closely. The total projected male enrollment as given by the last column is 292,300 for 1983-84, which is only 2,000 different than that projected by the grade specific time series.

Projection by time series has obvious drawbacks. Changing patterns of school enrollment caused by changes in births, deaths, migration, and educational policy cannot be accounted for by such a method. The effects of this are most noticeable over the long term. Over the short term time series methods work surprisingly well. The projected enrollment by time series for 1970-71 was the closest of all methods to the actual enrollment.

In an attempt to overcome some of the previously mentioned drawbacks we look at more complicated forms of Markov models in the next two chapters.

CHAPTER VI

THE MARKOV CHAIN MODEL

6.1 INTRODUCTION

In this chapter we consider a Markov chain model as a representation of part of the educational system. That is, the educational process, the flow of students through the system, is seen as a Markov process. In this way we think of the educational process as a series of events subject to random fluctuations and proceeding with time such that the development of the system depends only on its state at the present time and on inputs to the system, in the form of new entrants, at given points in time in the future.

The Markov chain model permits us to investigate the structure of the educational system and to project future enrollment distribution of students. Estimates of the time students can expect to spend at each level of the system given their present level and variance of these estimates can also be obtained, as well as graduate and dropout probabilities for each level. An advantage of the Markov chain model is that it may be readily updated and adjusted to allow revised input estimates to be incorporated into the model. These may arise from changing conditions or additional information.

In the present study the Markov model is applied to the Alberta educational system and tested using actual data. The model is then used to make projections and the theory of absorbing

Markov chains is applied to estimate grade times, the time spent in each grade, and graduate and dropout probabilities. The Markov chain model does not take the age factor into account, however, this has been done for the simulation in the next chapter.

6.2 THE MODEL

The Markov chain model is based on a matrix of transition probabilities, whose elements are the probabilities of passage between states within the system, and probabilities of entering and leaving the system for each state. As a representation of the educational system the states of the Markov chain are: the grades within the system, not entered school, and left school. The required probabilities are the probabilities of passage between grades, and of entering and leaving the system for each grade.

We assume, in the present case, that changes in the educational system, the progress of students, occur only at a specified time once a year. Therefore, the time variable is discrete and the transition probabilities only apply at given points in time. These times are taken as the beginning of each school year, the system is assumed to remain unchanged until the same time the following year. The development that follows coincides closely with that presented in Bartholomew (1967).

In addition to the assumptions outlined above, the following assumptions are also imposed on the model in the present study.

1. All new entrants enter grade one.
2. Once a students drops out of school he does not re-enter the system.
3. No student advance more than one grade at a time or are demoted.
4. The n^{th} step transition probabilities $p_{ij}(n)$ are invariant with time and do not depend on the number of steps n taken to attain state i.

(6.1)

The assumptions have been imposed to simplify the model in cases where a greater refinement would not contribute significantly to the accuracy of the model and would require additional parameter estimates that cannot be easily estimated with any degree of accuracy.

The first assumption implies all new students start in grade 1. Accurate migration estimates are not obtainable, and existing figures have shown very small positive or negative totals for recent years, therefore, the effect of migration has been ignored. The number of students returning to school does appear to be increasing, especially for the high school grades, however, these students can be accounted for in terms of reduced dropout rates. Although some students do advance more than one grade at a time and some students are enrolled in more than one grade at a time, these numbers tend to be small.

The last assumption implies the progress of a student in a given state does not depend on the number of steps taken to

attain that state. For example, the probability of a student repeating a grade does not change, no matter how many times he has previously repeated the grade.

We use the following definitions in the development of the model:

1. $S_j(0)$, $j = 1,2,\ldots,12$ is the initial grade size.
2. $S_j(t)$, $j = 1,2,\ldots,12$ is the number of students in grade j for time t to $t+1$.
3. $E(t)$ is the expected number of new entrants to the system at time t.
4. p_{ij} is the probability that a student advances from grade i to grade j.
5. $p_i = 1 - \sum_{j=1}^{12} p_{ij}$ is the probability of loss from grade i.
6. p_{0j} is the probability of a new entrant entering grade j.

(6.2)

Definitions (6.2) along with assumptions (6.1) imply,

$$p_{ij} = 0 \quad \text{if} \quad j \neq i \quad \text{and} \quad j \neq i+1 \ , \tag{6.3}$$

$$p_i = 1 - p_{ii} - p_{ii+1} \ , \tag{6.4}$$

$$\left.\begin{aligned} p_{01} &= 1 \\ p_{0j} &= 0 \quad \text{for} \quad j = 2,3,\ldots,12. \end{aligned}\right\} \qquad (6.5)$$

Applying these results, the probabilities that specify the Markov process may be tables as follows:

FIGURE 6.1

Probabilities for the Markov Model

Grade	1	2	3	...	11	12	
0	1	0	0	...	0	0	
1	p_{11}	p_{12}	0	...	0	0	p_1
2	0	p_{22}	p_{23}	...	0	0	p_2
3	0	0	p_{33}	...	0	0	p_3
.	.	.	.	...	.	.	.
.	.	.	.	...	.	.	.
.	.	.	.	...	.	.	.
11	0	0	0	...	$p_{11,11}$	$p_{11,12}$	p_{11}
12	0	0	0	...	0	$p_{12,12}$	p_{12}

The first row of Figure 6.1 represents the entrance probabilities for the system, the last column the dropout probabilities except p_{12} which includes graduates, and the 12 by 12 matrix is the matrix of transition probabilities for

the 12 grades.

If we denote the transpose of the matrix of transition probabilities by $\underline{P}'$ and the vector of new entrants by $\underline{E}(t)$ the expected grade size $\bar{S}_j(t)$ at time t is given by

$$\bar{S}_j(t) = \sum_{i=1}^{12} p_{ij}\, \bar{S}_i(t-1) + E(t)p_{0j} \quad , \qquad (6.6)$$

$$(j = 1,2,\ldots,12; \quad t = 1,2,3,\ldots)$$

or in vector notation

$$\underline{\bar{S}}(t) = \underline{P}'\, \underline{\bar{S}}(t-1) + E(t)\underline{p}_0 \quad , \qquad (6.7)$$

where $\underline{P}_0' = [P_{01},0,\ldots,0]$.

Equation (6.7), our basic equation in the application of the Markov chain model to the educational system, expresses the expected grade distribution at time t as the transformed grade totals for the previous year $t-1$ plus the new entrants to the system for the year t.

6.3 THE DATA AND PARAMETER ESTIMATION

In the present study a Markov chain model is applied to the school enrollment data for the Alberta educational system. This data is used to obtain estimates of the transition probabilities and new entrant totals for the model, as well as to test the adequacy of the model in predicting grade enrollment and annual enrollment totals. Transition probabilities are estimated separately for male and female categories as well as for the two

categories combined.

To apply the model equation (6.7), we must first obtain the necessary probability and new entrant estimates. As it was not possible to obtain all the required estimates directly from available data this was done on a stepwise basis. The first step was to derive such estimates as were possible from the data contained in the Department of Education Annual Reports. From the totals given in the tables of these reports, it was possible to obtain initial estimates of grade 1 and grade 9 failure rates as well as dropout rates for grades 7 to 12. These estimates are listed in Figures 6.2 and 6.3.

FIGURE 6.2

Grade 1 and Grade IX Failure Rates

	Grade I			Grade IX
Year	Male	Female	Combined	Combined
1960	.080	.043	.062	.101
1961	.078	.040	.060	.101
1962	.072	.039	.057	.099
1963	.073	.042	.058	.098
1964	.077	.046	.062	.097
1965	.073	.044	.059	.094
1966	.074	.046	.061	.090
1967	.076	.048	.063	.095
1968	.074	.045	.060	.086

FIGURE 6.3

Dropout Rates

	Grade						
	< 7	7	8	9	10	11	12
1965							
Male	.016	.012	.028	.065	.074	.116	.305
Female	.012	.008	.018	.046	.053	.106	.292
Combined	.015	.010	.023	.056	.064	.111	.299
1966							
Male	.006	.012	.026	.055	.072	.111	.334
Female	.004	.007	.015	.042	.047	.089	.370
Combined	.005	.010	.021	.049	.059	.100	.350
1967							
Male	.005	.011	.023	.053	.077	.121	.383
Female	.004	.006	.013	.032	.057	.080	.389
Combined	.004	.008	.018	.042	.067	.101	.386
1968							
Male	.006	.009	.021	.052	.082	.113	.378
Female	.004	.005	.011	.029	.055	.093	.408
Combined	.005	.007	.016	.041	.069	.103	.392

Additional estimates required were taken from other available sources, such as Li (1970), and a school dropout study obtained from the Department of Youth and Education, Province of Manitoba.

Once these first estimates had been made equation (6.7) was used to predict the known annual enrollment figures for the school years 1962 to 1970, using the 1961 enrollment figures as the initial grade sizes.

To put this more precisely the initial grade sizes $S_j(0)$, $(j = 1,2,\ldots,12)$ were taken as the 1961 enrollment figures for grades 1 to 12 and the expected number of new entrants $E(t)$, for the nine years to be predicted, were taken as those reported in the Department of Education Annual Reports except for the years 1969 and 1970. New entrant totals were not available for these latter years, therefore, the totals were estimated on the basis of the previous years totals. The transposed matrix of transition probabilities $\underline{P}'$ comprised the estimates obtained from available sources as described above.

Solutions of equation (6.7) were obtained from the individual and combined male and female categories using a computer program and the expected enrollment totals so derived were compared to the actual enrollments by calculating the error of each estimated grade enrollment total as a percentage of the actual grade enrollment total for each of the school years 1962 to 1970. This being done, adjustments where necessary, were made by hand to the estimated transition probabilities to increase the accuracy of prediction. This procedure was repeated a number of times until a satisfactory

FIGURE 6.4

Transition Matrices

Combined

Grade	1	2	3	4	5	6	7	8	9	10	11	12
1	0.06	0.94	0	0	0	0	0	0	0	0	0	0
2	0	0.045	0.95	0	0	0	0	0	0	0	0	0
3	0	0	0.035	0.96	0	0	0	0	0	0	0	0
4	0	0	0	0.03	0.96	0	0	0	0	0	0	0
5	0	0	0	0	0.035	0.96	0	0	0	0	0	0
6	0	0	0	0	0	0.04	0.95	0	0	0	0	0
7	0	0	0	0	0	0	0.055	0.93	0	0	0	0
8	0	0	0	0	0	0	0	0.05	0.92	0	0	0
9	0	0	0	0	0	0	0	0	0.07	0.85	0	0
10	0	0	0	0	0	0	0	0	0	0.06	0.87	0
11	0	0	0	0	0	0	0	0	0	0	0.03	0.03
12	0	0	0	0	0	0	0	0	0	0	0	0.23

Males

Grade	1	2	3	4	5	6	7	8	9	10	11	12
1	0.075	0.92	0	0	0	0	0	0	0	0	0	0
2	0	0.055	0.94	0	0	0	0	0	0	0	0	0
3	0	0	0.04	0.95	0	0	0	0	0	0	0	0
4	0	0	0	0.04	0.95	0	0	0	0	0	0	0
5	0	0	0	0	0.04	0.95	0	0	0	0	0	0
6	0	0	0	0	0	0.03	0.96	0	0	0	0	0
7	0	0	0	0	0	0	0.055	0.92	0	0	0	0
8	0	0	0	0	0	0	0	0.04	0.92	0	0	0
9	0	0	0	0	0	0	0	0	0.055	0.8	0	0
10	0	0	0	0	0	0	0	0	0	0.06	0.85	0
11	0	0	0	0	0	0	0	0	0	0	0.03	0.81
12	0	0	0	0	0	0	0	0	0	0	0	0.2

Females

Grade	1	2	3	4	5	6	7	8	9	10	11	12
1	0.045	0.95	0	0	0	0	0	0	0	0	0	0
2	0	0.035	0.96	0	0	0	0	0	0	0	0	0
3	0	0	0.025	0.97	0	0	0	0	0	0	0	0
4	0	0	0	0.03	0.96	0	0	0	0	0	0	0
5	0	0	0	0	0.03	0.96	0	0	0	0	0	0
6	0	0	0	0	0	0.035	0.95	0	0	0	0	0
7	0	0	0	0	0	0	0.055	0.93	0	0	0	0
8	0	0	0	0	0	0	0	0.05	0.93	0	0	0
9	0	0	0	0	0	0	0	0	0.03	0.92	0	0
10	0	0	0	0	0	0	0	0	0	0.06	0.88	0
11	0	0	0	0	0	0	0	0	0	0	0.03	0.9
12	0	0	0	0	0	0	0	0	0	0	0	0.2

reduction in the error of estimates was obtained. Figure 6.4 gives the estimates derived for combined, male, and female categories.

The rows and columns of the transition matrices in Figure 6.4 represent the grades 1 to 12 with the diagonal elements p_{ii} estimates of the probabilities of repeating and the upper diagonal elements $p_{i,1+1}$ estimates of the probabilities of passing each grade. The row sums subtracted from one give the dropout estimates except for grade 12 where the difference includes the probability of graduation, which must be estimated separately. The probabilities for each grade are tabled below.

FIGURE 6.5

Graduate and Dropout Probabilities

Grade	Combined		Male		Female	
	G	D	G	D	G	D
1	0	0	0	0	0	0
2	0	0.005	0	0.005	0	0.005
3	0	0.005	0	0.005	0	0.005
4	0	0.005	0	0.005	0	0.005
5	0	0.005	0	0.005	0	0.005
6	0	0.005	0	0.005	0	0.01
7	0	0.015	0	0.02	0	0.01
8	0	0.03	0	0.035	0	0.02
9	0	0.08	0	0.14	0	0.1
10	0	0.07	0	0.09	0	0.1
11	0	0.08	0	0.1	0	0.11
12	0.57	0.2	0.56	0.15	0.55	0.25

The estimated grade and total enrollment figures and the percentage error of estimate, using equation (6.7) with the transition matrices of Figure 6.4 for the school years 1961-62 to 1969-70 are given below. A negative percentage indicates an overestimation and a positive percentage an underestimation of the actual enrollment total.

Comparisons of estimated and actual enrollment totals appearing in Figures 6.6, 6.7, and 6.8 for school years 1961-62 to 1969-70 tend to confirm the validity of the assumption of constant transition probabilities for grades 1 to 9. The percentage error of grade estimate is quite small and fairly random for these grades. However, for the high school grades a definite increase in enrollment can be seen, beyond that which might be expected as a result of the natural increase in student population. This increase is shown to have taken place in the most part over the last four school years beginning with the grade 10 enrollment totals and extending to the grade 11 and 12 enrollment totals as the students progress through these grades. The effect takes place for grade 12 in school year 1968-69, as indicated by the 7.6 per cent overestimation for the previous year followed by a 1.1 per cent underestimation for this year. Reasons for this are probably a decrease in the number of students dropping out of high school and an increased number returning to complete their education after previously dropping out of school. These, in turn, could be due to such factors as a desire for further education, a demand for greater educational requirements by employers or a scarcity of employment opportunity.

FIGURE 6.6

ESTIMATED GRADE TOTALS: COMBINED

0	1962	1963	1964	1965	1966	1967	1968	1969	1970
1	35673	35322	36483	37262	38135	38338	38530	38812	39223
2	33979	35057	34761	35859	36640	37495	37725	37916	38189
3	31239	33278	34469	34248	35265	36042	36882	37130	37320
4	29977	30889	32873	34077	33900	34872	35647	36476	36739
5	29128	29912	30854	32803	34032	33905	34838	35618	36446
6	27028	28084	29839	30814	32723	33980	33908	34801	35586
7	26301	27258	28319	30054	31080	32960	34263	34267	35119
8	25816	25750	26638	27669	29334	30371	32171	33473	33542
9	24620	25474	25474	26290	27295	28898	29964	31695	33014
10	19459	22094	22978	23031	23728	24625	26041	27032	28563
11	15266	18304	20869	21869	22005	22624	23460	24767	25747
12	14679	16047	18883	21665	23134	23585	24203	25038	26315

ESTIMATED TOTAL ENROLMENT

1962	1963	1964	1965	1966	1967	1968	1969	1970
312063	327469	342461	355640	367273	377695	387632	397025	405808

PERCENT ERROR IN GRADE ESTIMATE

0	1	2	3	4	5	6	7	8	9	10	11	12
62	-0.33	-0.34	-0.25	-0.38	-0.6	-2.6	0.27	-2	-0.52	-6.9	-5.2	-3.7
63	-0.19	-0.43	-0.69	1.7	0.23	-1.1	-0.5	-0.84	-1.5	-6.2	-10	-9.2
64	-0.19	-0.8	-0.73	0.14	-1.5	-0.77	0.58	-1.6	-0.61	-6.9	-8.1	-13
65	-0.056	-0.29	-2.1	-0.33	-0.18	-0.38	0.43	-0.34	-0.98	-4.1	-8.3	-7.4
66	0.066	-0.36	-1.6	-0.74	-0.75	-1.4	1.3	-0.74	1.2	-4.5	-8	-6.2
67	0.27	-0.27	-0.98	0.5	-1.9	-2.3	-0.29	-1.2	-0.97	-0.086	-7.9	-7.4
68	0.051	-1.4	-0.46	1.2	-1.2	-1.5	-0.071	-0.51	-1.6	-4.2	-2.8	-7.6
69	-0.46	-0.024	1	-1.4	-1.2	-0.75	-0.27	-0.47	-0.58	-5.2	-4.3	-1.1
70	-1.1	-0.31	-1.2	-1.7	-1.8	-1.5	-0.18	-1.1	-0.86	-8.1	-4	-3.6

PERCENT ERROR IN YEARLY ESTIMATES

62	63	64	65	66	67	68	69	70
-1.4	-1.6	-1.7	-1.3	-1.4	-1.3	-0.43	1.1	1.9

FIGURE 6.7

ESTIMATED GRADE TOTALS:MALES

0	1962	1963	1964	1965	1966	1967	1968	1969	1970
1	18783	18456	18943	19425	19975	19968	19930	19995	20200
2	17755	18350	18081	18517	18986	19521	19544	19510	19568
3	16230	17339	17943	17714	19114	18572	19093	19135	19105
4	15569	16122	17204	17824	17629	18004	18456	18972	19033
5	14478	15447	16015	17070	17704	17544	17896	18041	18852
6	13717	14238	15179	15749	16775	17411	17277	17669	18044
7	13772	13995	14509	15446	16048	17070	17740	17648	17963
8	13205	13267	13476	13960	14846	15438	16407	17066	17227
9	12287	12889	12981	13179	13638	14483	15077	16006	16667
10	9200.8	10366	10933	11041	11296	11583	12281	12798	13573
11	7378.7	8484.8	9574.3	10155	10299	10452	10786	11410	11905
12	7846.7	8252.3	9265.8	10442	11254	11606	11832	12168	12775

ESTIMATED TOTAL ENROLMENT

1962	1963	1964	1965	1966	1967	1968	1969	1970
160202	167206	174104	180522	186474	191651	196319	200658	204687

PERCENT ERROR IN GRADE ESTIMATE

0	1	2	3	4	5	6	7	8	9	10	11	12
62	-0.27	-0.4	0.0074	-1.2	-0.45	-1.4	-0.12	-1.5	-0.97	-2.4	-1.7	-0.72
63	-0.19	-0.19	-0.44	-1.1	-0.23	-0.3	0.045	-1.1	-2	-1.7	-3.1	-2.6
64	-0.17	-0.28	-0.22	-0.72	-0.81	-0.55	-1.2	-0.97	-0.93	-2.1	-0.17	-3.1
65	-0.15	1.3	-1.3	-0.49	-0.6	-0.79	-0.23	-0.43	0.4	1.3	-0.15	2.6
66	-0.1	-0.11	-0.75	-0.97	-0.59	-1.4	1.3	-1	3	0.49	-0.036	3.5
67	-0.13	-0.28	-0.45	0.7	-1.5	-1.2	0.62	-1.3	-1.2	6.7	0.87	1.3
68	-0.048	2	-0.25	1.7	-0.79	-0.31	0.67	-1.5	-2	9.7	5.7	1.2
69	0.49	0.052	1.3	0.92	2.2	2.1	0.55	-1.8	-0.13	11	13	11
70	2.3	0.89	2.2	1.4	2.7	3.9	0.84	3	2.2	15	13	11

PERCENT ERROR IN YEARLY ESTIMATES

62	63	64	65	66	67	68	69	70
-0.81	-0.59	-0.42	0.24	0.17	0.37	1.3	3.1	4.4

FIGURE 6.8

ESTIMATED GRADE TOTALS:FEMALES

0	1962	1963	1964	1965	1966	1967	1968	1969	1970
1	16919	16896	17565	17859	18185	18398	18626	18838	19048
2	16109	16721	16721	17360	17663	17985	18200	18425	18635
3	14927	15838	16449	16464	17077	17384	17700	17914	18136
4	14622	14913	15810	16429	16463	17059	17374	17690	17903
5	13508	14516	14831	15702	16325	16376	16953	17275	17581
6	13109	13494	14480	14819	15671	16303	16374	16933	17263
7	12663	13216	13614	14577	14953	15788	16437	16541	17081
8	12479	12464	12980	13378	14298	14696	15497	16144	16273
9	12486	12605	12600	13079	13488	14377	14818	15597	16261
10	9387.6	10801	10984	10991	11384	11743	12493	12900	13564
11	7488	8559.1	9943.6	10112	10143	10476	10807	11467	11868
12	6747.4	7336.7	8314.7	9537.8	9997.5	10114	10403	10726	11319

ESTIMATED TOTAL ENROLMENT

1962	1963	1964	1965	1966	1967	1968	1969	1970
150441	157365	164192	170308	175649	180698	185682	190450	194944

PERCENT ERROR IN GRADE ESTIMATE

0	1	2	3	4	5	6	7	8	9	10	11	12
62	-0.57	-0.18	-0.016	-0.95	0.29	-2.2	-0.38	-1.4	-1.2	-1.9	-3.2	-5.9
63	-0.36	-0.77	-0.34	1.3	0.36	0.0054	-0.71	-0.42	-1.1	-1.8	-2.3	-10
64	0.078	-1.5	-0.8	0.18	2.3	0.25	1.4	-0.83	-1.1	-1.9	-0.9	-7.9
65	-0.081	-0.9	-2.4	-0.14	-0.47	1.6	1.3	1.3	-2.2	-0.54	-0.81	-0.95
66	0.11	-0.94	-2.1	-1.7	-0.91	0.36	1.9	0.86	0.5	0.44	-0.63	1.1
67	0.26	-0.31	-1	-0.86	-2.4	-1.9	0.58	0.52	1	4	-0.54	0.95
68	0.022	-0.76	-0.17	-0.36	-1.8	-1.2	-0.36	1.2	-0.71	7.9	5	1
69	0.32	-0.21	1.2	0.96	0.18	0.87	-0.68	0.72	-0.46	9.2	10	8.5
70	-0.29	-1.7	0.51	0.94	0.81	0.52	-0.068	0.65	-0.021	10	10.	12

PERCENT ERROR IN YEARLY ESTIMATES

62	63	64	65	66	67	68	69	70
-1.1	-0.87	-0.56	-0.21	-0.16	-0.095	0.71	2.1	2.4

The determination of such factors could be an area for further investigation in the study of student enrollment trends. The results might then be applied to the model in terms of revised parameter estimates which would increase the accuracy of prediction.

6.4 ABSORBING MARKOV CHAINS

A finite Markov chain is by definition a finite Markov process for which the transition probabilities $p_{ij}(n)$ do not depend on n, which is the case under assumptions for the present study. The states of a Markov chain may be divided into sets of transient and ergodic states. The set of transient states consists of those states which once left are never re-entered and the set of ergodic states consists of those states which once entered are never left. In the present context the transient states are the grades and the ergodic states are graduated and dropped out of school.

Consider the matrix of probabilities

$$\underline{T} = \left[\begin{array}{c|c} \underline{0} & \underline{I} \\ \hline \underline{Q} & \underline{R} \end{array}\right],$$

in which the identity matrix $\underline{I}$ corresponds to the set of ergodic states, the probability matrix $\underline{Q}$ corresponds to the set of transient states, $\underline{0}$ is a matrix of zeros and $\underline{R}$ is the matrix probabilities concerning the transition from transient to ergodic states. For the present study $\underline{T}$ has the form shown in Figure 6.9.

FIGURE 6.9

Probability Matrix

$$\underline{T} = \begin{array}{c|cccccc|cc} & 1 & 2 & \cdots & 11 & 12 & G & D \\ G & 0 & 0 & \cdots & 0 & 0 & 1 & 0 \\ D & 0 & 0 & \cdots & 0 & 0 & 0 & 1 \\ \hline 1 & P_{11} & P_{12} & \cdots & 0 & 0 & 0 & P_1 \\ 2 & 0 & P_{22} & \cdots & 0 & 0 & 0 & P_2 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots \\ 11 & 0 & 0 & \cdots & P_{11,11} & P_{11,12} & 0 & P_{11} \\ 12 & 0 & 0 & \cdots & 0 & P_{12,12} & P'_{12} & P_{12} \end{array}$$

14×14

where the elements of $\underline{T}$ have the definitions given in section 6.2 with p'_{12} the probability of graduation, $\underline{Q}$ the matrix of transition probabilities $\underline{P}$ for the grades 1 to 12, and $\underline{I}$ the matrix of transition probabilities for the states G (graduated) and D (dropped out of school).

If we define,

$$\left.\begin{aligned} \underline{N} &= (\underline{I} - \underline{p})^{-1} \\ N_2 &= \underline{N}(2N_{dg} - 1) - \underline{N}_{sq} \\ B &= \underline{N}\,\underline{R} \\ \underline{\tau} &= \underline{N}\,\underline{\xi} \\ \underline{\tau}_2 &= (2\underline{N} - \underline{I})\underline{\tau} - \underline{\tau}_{sq} \end{aligned}\right\} \qquad (6.8)$$

where,

(i) $\underline{I}$ is an identity matrix with the dimensions of P,

(ii) $\underline{N}_{dg}$ is the matrix formed by setting all off diagonal elements of $\underline{N}$ to zero,

(iii) $\underline{N}_{sq}$ is the matrix of squared elements of $\underline{N}$,

(iv) $\underline{\xi}$ is a column vector with elements 1, then it can be shown (see Kemeny and Snell (1969)) that

(i) $\underline{N}$ = the matrix of mean state time for the process,

(ii) $\underline{N}_2$ = the variance matrix for the mean times of $\underline{N}$,

(iii) $\underline{B}$ = the matrix of probabilities whose ij^{th} element, b_{ij}, is the probability that the process starting in transient state S_i ends up in absorbing state S_j.

(iv) $\underline{\tau}$ = the vector of mean process times for each state S_j.

(v) $\underline{\tau}_2$ = the vector of variances for the mean process times of $\underline{\tau}$.

For the educational process the preceding quantities have the following interpretation:

(i) The element n_{ij} of $\underline{N}$ is the time a student in grade i can expect to spend in grade j.

(ii) The element b_{ij} of $\underline{B}$ is the probability that a student in grade i will end up in absorbing state j, where state j may be graduated or dropped out of school.

(iii) The element τ_j of $\underline{\tau}$ is the length of time a student in grade j can be expected to spend in school.

The matrix of estimated transition probabilities $\underline{P}$ and dropout and graduate probabilities $\underline{R}$ obtained in section 6.3 were applied to equations (6.8) to obtain process estimates. The estimates are given in Figures 6.10, 6.11, and 6.12.

The rows and columns of the mean grade times and variance matrices corresponds to the grades 1 to 12. The $(i,j)^{th}$ element of each mean grade times matrix $\underline{N}$ is the time a student in grade i can expect to spend in grade j. In the combined case of Figure 6.10, for example, the expected time a grade 9 student will spend in grade 11 is .93 years and a grade 12 student can expect to spend 1.3 years in school before he graduates or drops out of school. Grade time estimates less

FIGURE 6.10

PROCESS ESTIMATES:COMBINED

N=MATRIX OF MEAN GRADE TIMES

1	2	3	4	5	6	7	8	9	10	11	12
1.1	1	1	1	1	1	1	1	1	0.9	0.86	0.93
0	1	1	1	1	1	1	1	1	0.9	0.86	0.93
0	0	1	1	1	1	1	1	1	0.91	0.87	0.94
0	0	0	1	1	1	1	1	1	0.91	0.87	0.94
0	0	0	0	1	1	1	1	1	0.92	0.88	0.95
0	0	0	0	0	1	1	1	1	0.92	0.88	0.95
0	0	0	0	0	0	1.1	1	1	0.93	0.89	0.95
0	0	0	0	0	0	0	1	1	0.94	0.9	0.97
0	0	0	0	0	0	0	0	1.1	0.97	0.93	1
0	0	0	0	0	0	0	0	0	1.1	1	1.1
0	0	0	0	0	0	0	0	0	0	1.1	1.2
0	0	0	0	0	0	0	0	0	0	0	1.3

VARIANCE MATRIX FOR N

1	2	3	4	5	6	7	8	9	10	11	12
0.07	0.05	0.04	0.04	0.05	0.06	0.09	0.1	0.15	0.2	0.29	0.62
0	0.05	0.04	0.04	0.05	0.06	0.09	0.1	0.15	0.2	0.29	0.62
0	0	0.04	0.04	0.05	0.06	0.08	0.09	0.15	0.2	0.29	0.62
0	0	0	0.03	0.04	0.05	0.08	0.09	0.14	0.2	0.28	0.62
0	0	0	0	0.04	0.05	0.07	0.08	0.14	0.19	0.28	0.62
0	0	0	0	0	0.04	0.07	0.08	0.13	0.19	0.28	0.62
0	0	0	0	0	0	0.06	0.07	0.13	0.19	0.28	0.61
0	0	0	0	0	0	0	0.06	0.11	0.18	0.27	0.61
0	0	0	0	0	0	0	0	0.08	0.15	0.25	0.6
0	0	0	0	0	0	0	0	0	0.07	0.18	0.55
0	0	0	0	0	0	0	0	0	0	0.11	0.49
0	0	0	0	0	0	0	0	0	0	0	0.33

T=MEAN PROCESS TIMES

1	2	3	4	5	6	7	8	9	10	11	12
12	11	9.9	8.9	7.9	6.9	5.9	4.9	4	3.2	2.3	1.3

VARIANCE MATRIX FOR T

1	2	3	4	5	6	7	8	9	10	11	12
4.5	4.4	3.9	3.5	3.1	2.9	2.7	2.3	1.8	0.98	0.6	0.39

GRADUATE AND DROPOUT PROBABILITIES BY GRADE

1	0.53	0.47
2	0.53	0.47
3	0.53	0.47
4	0.54	0.46
5	0.54	0.46
6	0.54	0.46
7	0.54	0.46
8	0.55	0.45
9	0.57	0.43
10	0.62	0.38
11	0.68	0.32
12	0.74	0.26

FIGURE 6.11

PROCESS ESTIMATES:MALE

N=MATRIX OF MEAN GRADE TIMES

1	2	3	4	5	6	7	8	9	10	11	12
1.1	1.1	1	1	1	1	1	0.99	0.97	0.83	0.77	0.88
0	1.1	1	1	1	1	1	0.99	0.97	0.83	0.77	0.88
0	0	1	1	1	1	1	1	0.98	0.83	0.78	0.89
0	0	0	1	1	1	1	1	0.98	0.84	0.78	0.89
0	0	0	0	1	1	1	1	0.99	0.84	0.79	0.9
0	0	0	0	0	1	1	1	0.99	0.85	0.79	0.9
0	0	0	0	0	0	1.1	1	1	0.85	0.79	0.91
0	0	0	0	0	0	0	1	1	0.87	0.81	0.92
0	0	0	0	0	0	0	0	1.1	0.9	0.84	0.96
0	0	0	0	0	0	0	0	0	1.1	0.99	1.1
0	0	0	0	0	0	0	0	0	0	1.1	1.2
0	0	0	0	0	0	0	0	0	0	0	1.4

VARIANCE MATRIX FOR N

1	2	3	4	5	6	7	8	9	10	11	12
0.09	0.06	0.05	0.05	0.06	0.05	0.09	0.09	0.14	0.25	0.33	0.82
0	0.06	0.05	0.05	0.06	0.05	0.09	0.09	0.14	0.25	0.33	0.82
0	0	0.04	0.05	0.05	0.05	0.08	0.08	0.14	0.25	0.33	0.82
0	0	0	0.04	0.05	0.04	0.06	0.08	0.13	0.24	0.33	0.82
0	0	0	0	0.04	0.04	0.07	0.07	0.13	0.24	0.32	0.83
0	0	0	0	0	0.03	0.07	0.07	0.12	0.24	0.32	0.83
0	0	0	0	0	0	0.06	0.06	0.12	0.24	0.32	0.83
0	0	0	0	0	0	0	0.04	0.1	0.23	0.31	0.83
0	0	0	0	0	0	0	0	0.06	0.2	0.3	0.82
0	0	0	0	0	0	0	0	0	0.07	0.2	0.77
0	0	0	0	0	0	0	0	0	0	0.11	0.71
0	0	0	0	0	0	0	0	0	0	0	0.58

T=MEAN PROCESS TIMES

1	2	3	4	5	6	7	8	9	10	11	12
12	11	9.6	8.6	7.6	6.6	5.6	4.7	3.8	3.2	2.4	1.4

VARIANCE MATRIX FOR T

1	2	3	4	5	6	7	8	9	10	11	12
5.1	5.1	4.5	4.1	3.8	3.5	3.4	2.9	2.5	1.3	0.81	0.58

GRADUATE AND DROPOUT PROBABILITIES BY GRADE

1	0.49	0.51
2	0.49	0.51
3	0.5	0.5
4	0.5	0.5
5	0.5	0.5
6	0.5	0.5
7	0.51	0.49
8	0.52	0.48
9	0.54	0.46
10	0.63	0.37
11	0.7	0.3
12	0.79	0.21

FIGURE 6.12

PROCESS ESTIMATES:FORMAL?

N=MATRIX OF MEAN GRADE TIMES

1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	1	1	1	1	1	0.89	0.82	0.82
0	1	1	1	1	1	1	1	1	0.89	0.82	0.82
0	0	1	1	1	1	1	1	1	0.89	0.83	0.83
0	0	0	1	1	1	1	1	1	0.9	0.83	0.83
0	0	0	0	1	1	1	1	1	0.9	0.83	0.83
0	0	0	0	0	1	1	1	1	0.91	0.84	0.84
0	0	0	0	0	0	1.1	1	1	0.92	0.85	0.85
0	0	0	0	0	0	0	1	1.1	0.93	0.86	0.86
0	0	0	0	0	0	0	0	1.1	0.95	0.88	0.88
0	0	0	0	0	0	0	0	0	1.1	0.98	0.98
0	0	0	0	0	0	0	0	0	0	1.1	1.1
										0	1.2

VARIANCE MATRIX FOR N

1	2	3	4	5	6	7	8	9	10	11	12
0.05	0.04	0.03	0.04	0.05	0.06	0.09	0.1	0.16	0.21	0.31	0.56
0	0.04	0.03	0.04	0.05	0.06	0.09	0.1	0.16	0.21	0.31	0.56
0	0	0.03	0.04	0.04	0.05	0.09	0.09	0.15	0.21	0.31	0.56
0	0	0	0.03	0.04	0.05	0.08	0.09	0.15	0.21	0.31	0.56
0	0	0	0	0.03	0.04	0.08	0.08	0.15	0.2	0.3	0.55
0	0	0	0	0	0.04	0.07	0.08	0.14	0.2	0.3	0.55
0	0	0	0	0	0	0.05	0.07	0.13	0.19	0.3	0.55
0	0	0	0	0	0	0	0.06	0.12	0.19	0.29	0.55
0	0	0	0	0	0	0	0	0.09	0.17	0.26	0.55
0	0	0	0	0	0	0	0	0	0.07	0.21	0.51
0	0	0	0	0	0	0	0	0	0	0.11	0.44
0	0	0	0	0	0	0	0	0	0	0	0.31

T=MEAN PROCESS TIMES

1	2	3	4	5	6	7	8	9	10	11	12
12	11	9.7	8.7	7.7	6.7	5.8	4.8	3.8	3	2.2	1.2

VARIANCE MATRIX FOR T

1	2	3	4	5	6	7	8	9	10	11	12
4.4	4.3	3.8	3.4	3.1	2.8	2.5	2.2	1.9	1	0.55	0.31

GRADUATE AND DROPOUT PROBABILITIES BY GRADE

1	0.45	0.55
2	0.45	0.55
3	0.45	0.55
4	0.46	0.54
5	0.46	0.54
6	0.46	0.54
7	0.47	0.53
8	0.47	0.53
9	0.48	0.52
10	0.54	0.46
11	0.6	0.4
12	0.69	0.31

than one year are accounted for by the fact that some students will drop out of school before attaining the grade, and zero grade times reflect the assumption of no demotions.

Mean process times $\underline{\tau}$ give the future number of years a student starting a given grade can expect to spend in school. Thus for the combined case a student starting grade 4 can expect to spend 8.9 additional years in school.

Graduate and dropout probabilities are the corresponding probabilities for students having attained a given grade. Thus, for the combined case the estimated probability that a grade 10 student will graduate is .62.

6.5 PROJECTIONS

Projected enrollments for the school years 1970-71 to 2004-05 were obtained for males and females separately and combined using the estimated new entrant totals tabulated in Figure B.2.2. These projections were derived from equation (6.7) using the transpose of the transition matrices P' given in Figure 6.4, and the new entrant totals $E(t)$ in Figure B.2.2. The projected school enrollments are tabulated for the combined and separate male and female cases in Figures 6.13, 6.14, and 6.15, for the school years ending 1971 to 2005. The projected total enrollment figures are given in 100,000's, thus, for example, the projected total enrollment in Figure 6.13 for the combined case and school year 1970-71 is $4.1818 \times 100{,}000 = 418{,}180$. A graph of the projected total enrollment totals for the school years 1970-71 to 2004-05 is given in Figure 6.16.

FIGURE 6.13

Projected School Enrollment - Alberta Schools
1971 to 2005

Combined

Year	Grade												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1971	37691	38999	37490	37374	37191	37075	36261	34417	33530	30169	29449	28531	4.1818E5
1972	33531	37185	38361	37112	37181	37187	37215	35444	34010	30311	28897	31005	4.1744E5
1973	31793	33193	36668	37940	36929	37181	37374	36383	34989	30728	28971	31116	4.1326E5
1974	31169	31379	32816	36339	37715	36939	37378	36577	35921	31584	29340	31203	4.0839E5
1975	31162	30711	30959	32594	36206	37684	37148	36590	36166	32428	30119	31529	4.0337E5
1976	32566	30674	30259	30698	32557	36265	37843	36377	36194	32686	30923	32251	3.9929E5
1977	34901	31992	30200	29970	30610	32706	36533	37013	36000	32726	31220	33084	3.9695E5
1978	36739	34246	31450	29891	29842	30694	33080	35826	36572	32564	31282	33522	3.9571E5
1979	37958	36076	33635	31089	29740	29876	30978	32555	35520	33040	31146	33674	3.9529E5
1980	39118	37304	35449	33222	30886	29745	30086	30438	32437	32175	31548	33596	3.9675E5
1981	40270	38450	36680	35028	32974	30840	29913	29502	30273	29502	30831	33912	3.9818E5
1982	41426	39584	37811	36263	34781	32889	30943	29294	29261	27502	28442	33290	4.0159E5
1983	42660	40721	38928	37387	36030	34705	32946	30242	28999	26522	26487	31286	4.0601E5
1984	43957	41932	40048	38492	37152	35977	34782	32152	29853	26240	25458	29180	4.1529E5
1985	45262	43207	41237	39601	38253	37105	36091	33955	31670	26949	25120	27842	4.2628E5
1986	46571	44491	42490	40776	39356	38207	37235	35263	33455	28536	25707	27253	4.3934E5
1987	47888	45779	43753	42013	40522	39310	38345	36392	34784	30149	27140	27565	4.5368E5
1988	48955	47075	45022	43263	41751	40474	39453	37480	35915	31375	28672	28675	4.6831E5
1989	49834	48136	46297	44519	42994	41700	40620	38565	36996	32410	29877	30439	4.8230E5
1990	50696	49010	47349	45780	44243	42942	41849	39705	38070	33391	30886	31799	4.9572E5
1991	51555	49860	48217	46829	45498	44101	43097	40905	39193	34363	31830	32949	5.0849E5
1992	52415	50705	49055	47693	46548	45445	44351	42126	40376	35376	32760	33997	5.2085E5
1993	53258	51552	49887	48523	47415	46504	45612	43353	41582	36442	33726	35010	5.3286E5
1994	54087	52382	50721	49347	48242	47378	46688	44587	42796	37531	34740	36045	5.4454E5
1995	54916	53199	51538	50172	49062	48207	47577	45649	44016	38628	35779	37125	5.5587E5
1996	55744	54015	52343	50982	49882	49027	48414	46529	45078	39731	36827	38235	5.6681E5
1997	56571	54830	53146	51779	50688	49848	49239	47351	45962	40700	37881	39360	5.7736E5
1998	57397	55644	53949	52574	51482	50655	50064	48160	46781	41510	38818	40494	5.8753E5
1999	58223	56457	54750	53368	52273	51449	50876	48968	47581	42254	39607	41533	5.9734E5
2000	59050	57271	55551	54161	53063	52240	51675	49763	48381	42960	40326	42427	6.0688E5
2001	59876	58084	56351	54953	53852	53030	52470	50545	49168	43702	41021	43228	6.1628E5
2002	60703	58897	57152	55746	54640	53819	53265	51324	49944	44415	41713	43990	6.2561E5
2003	61529	59711	57953	56538	55428	54607	54058	52102	50714	45117	42395	44740	6.3489E5
2004	62355	60524	58754	57331	56217	55396	54850	52879	51484	45814	43067	45478	6.4415E5
2005	63182	61338	59554	58123	57005	56184	55643	53654	52252	46510	43734	46206	6.5339E5

FIGURE 6.14

Projected Enrollment - Alberta Schools

1971 to 2005

Males

Year	Grade												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1971	19697	20106	19341	19330	19120	18975	19026	17367	17067	14591	14802	15205	2.1463E5
1972	17557	19227	19673	19147	19128	18733	19263	18199	16916	14529	13735	16399	2.1251E5
1973	16481	17210	18860	19455	18955	18734	19043	18450	17673	14405	13586	15881	2.0873E5
1974	16262	16109	16932	18695	19241	18569	19032	18258	17846	15003	13467	15610	2.0512E5
1975	16277	15847	15819	16833	18530	18836	18873	18239	17784	15257	13964	15435	2.017E5
1976	16839	15847	15529	15702	16733	18169	19120	18093	17758	15143	14225	15787	1.9894E5
1977	18005	16364	15517	15381	15586	16441	18494	18315	17622	15115	14152	16101	1.9709E5
1978	18947	17465	16002	15356	15235	15300	16801	17747	17819	15004	14122	16132	1.9563E5
1979	19579	18392	17057	15817	15198	14932	15612	16167	17307	15155	14025	16117	1.9556E5
1980	20178	19024	17971	16837	15634	14886	15194	15010	15825	14755	14144	16034	1.9549E5
1981	20775	19610	18601	17746	16620	15299	15126	14579	14679	13545	13815	16107	1.961E5
1982	21371	20191	19178	18381	17523	16248	15519	14499	14220	12556	12757	15861	1.983E5
1983	22008	20772	19747	18954	18163	17135	16452	14857	14121	12129	11821	14933	2.0109E5
1984	22680	21390	20316	19518	18733	17769	17354	15730	14445	12025	11374	13905	2.0524E5
1985	23355	22042	20919	20081	19291	18329	18013	16595	15266	12278	11245	13245	2.1066E5
1986	24031	22699	21556	20676	19848	18876	18587	17235	16107	12950	11448	12949	2.1696E5
1987	24708	23357	22199	21305	20436	19422	19144	17789	16742	13663	12037	13028	2.2383E5
1988	25258	24016	22844	21942	21057	19997	19698	18324	17287	14214	12697	13529	2.3086E5
1989	25714	24559	23489	22579	21687	20604	20281	18855	17809	14682	13224	14208	2.3769E5
1990	26159	25007	24025	23217	22318	21221	20896	19412	18326	15128	13670	14832	2.4421E5
1991	26601	25442	24468	23752	22949	21838	21521	20001	18867	15569	14089	15374	2.5047E5
1992	27045	25872	24894	24195	23482	22457	22149	20599	19438	16028	14501	15871	2.5653E5
1993	27479	26305	25315	24617	23924	22982	22777	21201	20021	16512	14929	16349	2.6241E5
1994	27904	26727	25739	25034	24343	23417	23315	21803	20606	17007	15379	16834	2.6811E5
1995	28328	27142	26153	25453	24756	23828	23763	22322	21192	17505	15840	17339	2.7362E5
1996	28752	27555	26559	25864	25171	24233	24182	22755	21702	18004	16305	17859	2.7894E5
1997	29177	27968	26964	26266	25578	24640	24594	23158	22128	18442	16771	18386	2.8407E5
1998	29598	28381	27368	26666	25976	25038	25007	23553	22522	18809	17185	18916	2.8902E5
1999	30018	28791	27773	27067	26372	25428	25412	23948	22907	19146	17534	19406	2.936E5
2000	30438	29200	28175	27467	26768	25816	25809	24337	23292	19475	17852	19830	2.9846E5
2001	30858	29609	28575	27864	27164	26204	26203	24717	23671	19802	18160	20211	3.0304E5
2002	31278	30018	28975	28261	27558	26592	26597	25096	24042	20125	18466	20571	3.0758E5
2003	31698	30427	29376	28657	27950	26978	26991	25473	24410	20441	18768	20923	3.1209E5
2004	32118	30835	29776	29053	28342	27362	27383	25851	24778	20755	19064	21270	3.1659E5
2005	32537	31244	30176	29449	28734	27746	27774	26226	25145	21068	19357	21610	3.2107E5

FIGURE 6.15

Projected School Enrollment - Alberta Schools

1971 - 2005

Females

Year	Grade												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
1971	18020	19684	18051	18224	17887	17631	17423	16693	16534	14238	13874	13132	2.0039E5
1972	16001	17772	18388	18056	18032	17788	17708	17038	16847	14412	13208	13726	1.9898E5
1973	15337	15823	17521	18378	17875	17933	17873	17320	17193	14680	13295	13312	1.9654E5
1974	14926	15124	15628	17547	18179	17787	18020	17488	17483	14979	13527	13290	1.9399E5
1975	14906	14709	14910	15685	17390	18074	17889	17633	17662	15235	13800	13482	1.9138E5
1976	15749	14676	14493	14933	15580	17327	18155	17518	17811	15397	14039	13736	1.8942E5
1977	16913	15475	14451	14506	14803	15563	17460	17760	17717	15523	14197	13870	1.8835E5
1978	17809	16609	15217	14452	14370	14756	15745	17125	17934	15460	14322	14154	1.8795E5
1979	18398	17500	16325	15194	14305	14312	14884	15499	17361	15633	14275	14289	1.8798E5
1981	19514	18644	17797	17181	16090	14913	14315	14137	14858	13869	14044	14229	1.885E5
1982	20074	19191	18343	17778	16976	15968	14955	14020	14336	13016	12914	14113	1.8975E5
1983	20672	19742	18882	18326	17577	16856	15992	14609	14186	12536	12096	13154	1.9169E5
1984	21299	20330	19425	18866	18121	17463	16893	15603	14721	12384	11619	12307	1.9463E5
1985	21929	20945	20002	19408	18655	18007	17519	16490	15682	12815	11449	11737	1.9903E5
1986	22563	21565	20608	19984	19191	18539	18070	17118	16591	13634	11795	11519	2.0466E5
1987	23203	22190	21218	20589	19761	19072	18606	17661	17247	14423	12514	11738	2.1117E5
1988	23720	22820	21832	21199	20358	19638	19142	18186	17805	15008	13241	12359	2.1822E5
1989	24145	23333	22453	21813	20962	20231	19709	18711	18338	15500	13798	13065	2.2531E5
1990	24563	23755	22961	22434	21570	20832	20304	19265	18869	15967	14262	13651	2.3206E5
1991	24980	24166	23378	22945	22183	21436	20907	19846	19426	16430	14696	14140	2.3843E5
1992	25396	24576	23784	23365	22693	22045	21514	20435	20011	16915	15124	14585	2.4453E5
1993	25805	24986	24188	23771	23112	22557	22127	21030	20606	17424	15570	15016	2.5044E5
1994	26210	25389	24591	24175	23514	22977	22646	21630	21206	17942	16037	15459	2.5619E5
1995	26615	25788	24988	24579	23914	23377	23073	22142	21812	18466	16515	15921	2.6178E5
1996	27019	26186	25381	24976	24313	23775	23478	22565	22337	18994	16997	16396	2.6719E5
1997	27422	26585	25774	25369	24706	24173	23878	22962	22773	19456	17485	16877	2.7242E5
1998	27827	26981	26166	25761	25095	24564	24277	23355	23177	19841	17917	17363	2.7746E5
1999	28234	27380	26556	26154	25484	24951	24671	23746	23574	20195	18279	17806	2.8233E5
2000	28640	27780	26949	26544	25872	25338	25061	24132	23969	20542	18609	18184	2.8703E5
2001	29047	28181	27343	26937	26259	25724	25449	24513	24360	20888	18930	18524	2.9162E5
2002	29454	28581	27737	27331	26647	26109	25837	24893	24746	21228	19249	18849	2.9615E5
2003	29860	28981	28131	27725	27037	26495	26224	25274	25131	21565	19564	19169	3.0066E5
2004	30267	29382	28525	28119	27427	26883	26612	25652	25515	21901	19876	19485	3.0516E5
2005	30674	29782	28920	28513	27817	27271	27002	26032	25898	22236	20186	19798	3.1413E5

A steady drop in the projected total enrollment occurs up to school year 1981-82, due to the decline in the number of births for the years 1961 to 1969. The projected increase in the number of births for subsequent years has produced an annual increase in the projected enrollments for the school years 1981-82 to 2004-05, with a total enrollment of 653,390 students projected for 2004-05. The relationship between projected births and new entrants, is illustrated by Figure 6.16.

It may be noted that the sum of the projected enrollments for the individual male and female cases differs from the projected combined totals. This is primarily due to the independent estimation of the transition matrices for each of the three cases, with higher leaving rates occurring for the high school grades in the male and female cases. The result is, that the projected combined totals are generally higher than the sum of the individual totals for males and females. Present trends in high school enrollments indicate greater numbers of students are remaining in school or returning to school producing increased high school enrollments. Therefore, if these trends continue, we can expect the increased high school enrollment projected in the combined case.

The 1970-71 actual enrollments and projected enrollments using the Markov model are given in Figure 6.17.

FIGURE 6.16

Projected Total Enrollment for Alberta Schools
for School Years 1970-71 to 2004-05

(Unit = 10,000)

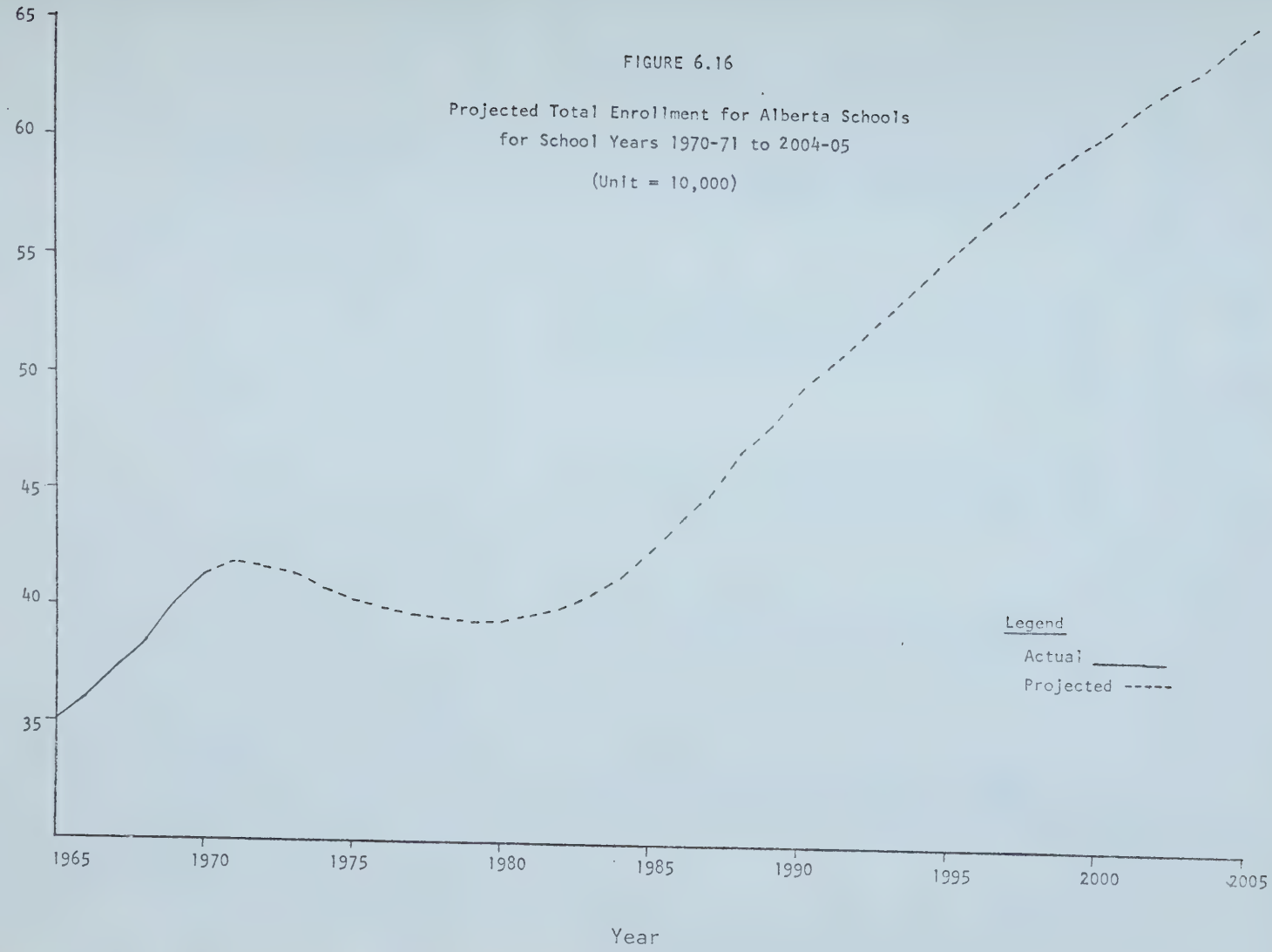


FIGURE 6.17

Actual and Projected Student Enrollment
School Year 1970-71

	Male		Female		Combined	
Grade	Actual	Projected	Actual	Projected	Actual	Projected
1	20308	19697	18918	18020	39226	37691
2	19799	20106	18642	18684	38443	38999
3	19458	19341	18557	18051	38015	37490
4	19444	19333	18406	18224	37850	37374
5	19382	19120	18283	17887	37665	37191
6	19173	18975	17998	17631	37171	37075
7	18994	19026	17581	17423	36575	36261
8	18054	17367	17120	16693	35174	34417
9	17323	17067	16459	16534	33782	33530
10	16738	14591	15785	14238	32523	30169
11	14615	14802	14092	13874	28707	29449
12	15348	15205	13445	13132	28793	28531

A comparison of the actual and projected enrollment totals for the school year 1970-71 indicates the projection is quite good except for grades 1 and 10, where the projected totals underestimate the actual totals to some degree. This indicates a greater number of students entering grade 1 than expected due, in part, to an increased number of five year olds entering grade one, and a greater number of students continuing on or returning to high school. A detailed investigation of these trends might produce additional information that could be used to adjust the parameter estimates for the model and increase the accuracy of projection.

CHAPTER VII

RANDOM WALK SIMULATION

7.1 INTRODUCTION

Simulation is a technique that involves formulating a model of a real situation and then performing experiments on the model. Simulation may take many forms but we restrict ourselves to a simulation based on a mathematical model. Churchman (1966) gives a formula definition of simulation as follows: "x simulates y" is true if and only if

(a) x and y are formal systems

(b) y is assumed to be the real system

(c) x is taken to be an approximation to the real system

(d) the rules of validity in x are non-error-free (the simulation can be wrong due to inadequate model, or approximating the model).

We take y to be that part of the formal educational system consisting of the enrollment of students in a graded system. x is our "simulation" of the real system y. It is a model based on information obtained by and from the development of the data base. Our primary goal consists of making x as close to being the same as y as we possibly can. Once we are reasonably satisfied with x we use x to make projections of student enrollments, our ultimate goal.

The first step in the analysis of our system is to isolate all the components of our system, and then formulate the logical and mathematical rules governing their interaction. In

Chapter 3 the variables of the student enrollment system: age, grade, year, and sex were separated out and discussed individually. In the same chapter the processes of the system; failure, passing, double promotion, dropout, graduation were classified in a closed form. The system was opened up by the introduction of the immigration and migration, the effect of which crosscuts all the variables in the closed system. The processes of the system represent risks which each student is exposed to in each year of his school career. Chapter 4 assisted us in determining rules of interaction and movement by the study of trends and tendency of the data base. Many of our rules are specified by previous empirical investigation. The description of each of the variables and their relationships are called the model of the system.

The purpose of our model is to abstract student flows in an education system and use this information to make projections of school enrollments. As part of this procedure we study how changes in certain parts of the model affect other aspects of our system.

7.2 ASSUMPTIONS

Although we try to model the real system as closely as possible we are forced into making assumptions about our model. Each assumption moves our approximate system further from the real system, thus limiting the validity of our results. Our students are "statistical". We generate and process these statistical students through our model individually. This alone means that

our model is not "true". The assumptions that are used simplify our model and we feel they are reasonable ones to use.

(1) The model disregards the mortality effect when age $a > 8$. It is possible to include mortality effect up to our limiting age but we found that the effect on our results was miniscule and the extra computer process time made it impractical.

(2) The student starting in the i^{th} grade-state S_i will move to another state S_j, $i \leq j$, $j = i, i+1, i+2$, independent of how he arrived at the S_i state. Therefore, the model is first order Markovian. We assume that a student starts off each grade with a clear record. If the student is in grade 4 we are not concerned about how he got to that grade.

(3) It is possible, in theory, for a student to fail any grade an infinite number of times. In our model the number of failures is restricted to the maximum number of years the student is in the system. However, it is extremely unlikely that this would occur.

(4) Students over the age of 18 will not return to the school system once they have left.

(5) Time is treated as a discrete variable.

(6) For the purposes of this model all transition and returning probabilities are assumed to be constant over time, i.e., the stationary nature of the Markov model is assumed. This means that the joint distribution of the sets of variables constituting the probabilities do not depend on the location of the time scale. In our discrete case any finite set of variables has the same joint

distribution as the time - shifted variables giving us a steady state operation.

(7) The number of births in each year and entering rates must be known or estimated for the future years that are to be simulated.

(8) Due to the nature of a random walk simulation the outputs may not be unique depending upon the initiating random number used in the random number generating proces, RANDU.

(9) The accuracy of the projections are dependent on the transition and re-entering parameters, and for short-range projections, the status tables of the initiating year.

7.3 DESCRIPTION OF THE MODEL

The simulation of an educational system by a random walk model involves a series of individual students, exposed to individual risks, being taken through a graded school system. In this way we simulate cross-sectional data in the form of age-grade tables determining the number of students in any year distributed jointly by age, grade, and sex. The purpose of this model is to project enrollment under various changes in parameters, exogenous variables, status variables, and operating characteristics of the system.

The component of the system is the student classified by age, grade, year, and sex variables where,

n - number of students

a - age of a student, $a = 0,1,\ldots,18^{+}$

(18^{+} includes students ages 18, 19, 20, 21)

g - grade or state of the student, $g = 1,2,\ldots,12$

y - school year, $y = 0,1,\ldots,24$

$\hat{y}$ - projected school years, $y = 25,26,\ldots,59$.

7.3.1 Exogenous Variables

Exogenous variables are input variables of the model. The variables have previously been calculated and are independent of the actual system being modelled.

Life Table

$$\underline{L} = \{L(a)\} \text{ - life table, a } 9 \times 1 \text{ vector, where } a = 0,1,\ldots,8. \qquad (7.1)$$

The life table shows the surviving probability of a subject a years old reaching age a+1. An assumption, stated earlier, indicates that we are not concerned with calculating the survival rate for students age 9 years and older. This sets the survival probability for each of these ages at 1.0. Because our table is not dependent on the year we are assuming that the life table is fixed and applies to any year.

Entering Probabilities

$$\underline{E}_y = \{E_y(a)\} \text{ - entering probabilities, a } 4 \times 1 \text{ vector, where } a = 5,6,7,8. \qquad (7.2)$$

The table $E_y(a)$ shows the probability of a subject a years old entering the school system in year y. We are assuming that no subject is allowed into the school system before the age of 5 and by the age of 9 all subjects have entered.

Migration Table

$$\underline{I}_y = \{I_y(a)\} \text{ - net migration in year } y, \text{ a } 1 \times 19 \text{ vector, where } a = 0,1,\ldots,18. \quad (7.3)$$

The net migration table shows the number of students entering (leaving) the province classified by age a in year y. If the table entry is positive we say that the total effect of migration to the province is that of immigration, i.e., more subjects have entered than left. A negative table entry indicates emigration.

The life table and the net migration table are uncontrollable exogenous variables. The policy makers of the system cannot manipulate or control them. They are generated by forces external to the educational system. The table of entering probabilities can be controlled by policy makers of the system as well as elected board members of individual school districts in the system.

7.3.2 Status Variables

Status variables describe the state of a system or one of its components either at the beginning or the end of our

accounting period. These variables interact with exogenous variables and parameters according to the relationship of the system. The value of any status variable in a certain year may depend on both the values of exogenous variables for some lag periods as well as the value of certain output variables in later time periods.

System Table

$$\underline{S}_y = \{S_y(g,a)\} \text{ - school enrollment (age-grade) table of grade by age in year } y, \text{ a } 12 \times 14 \text{ matrix, where } g = 1,2,\ldots,12 \quad a = 5,6,\ldots,18^+ . \tag{7.4}$$

The school enrollment table represents the number of students enrolled in grade g who are a years old in year y. This is our basic table showing the enrollment at any time. We employ this status variable as an input table of the base year. It also becomes an output table as projected values.

Waiting Table

$$\underline{W}_y = \{W_y(a)\} \text{ - pre-school table, a } 9 \times 1 \text{ vector, where } a = 0,1,\ldots,8. \tag{7.5}$$

$W_y(a)$ shows the number of pre-school children a years old in year y. When $a = 0$ we have the total number of births and in conjunction with the life table we can determine the number

of students surviving. By the age of 8 we assume that all subjects will have entered school. Due to government controls it has been shown that less than 1/2 % of eligible students fail to enter school. $\underline{I}_y$ and $\underline{W}_y$ together make up our new entrants into the system.

Dropout Table

$$\underline{D}_y = \{D_y(g,a)\} \text{ - dropout table, a } 12 \times 14 \text{ matrix,} \quad (7.6)$$

where, $g = 1,2,\ldots,12$

$a = 5,6,\ldots,18.$

The dropout table gives the number of students, a years old, whose last grade was g in year y, who have left the system. In reality the minimum school leaving age is 16, but we use the earlier ages to account for those who are forced into leaving school for health reasons or who fail to transfer to another system in the proper way.

7.3.3 Parameters

Parameters are also input variables in the model. These are actually exogenous variables that are estimated from external sources. For our purposes we use different levels of parameters to study the effects of the changes on the output variables of the system.

Transition Probabilities

$\underline{T} = \{T(a,g,i)\}$ - transition probabilities represent (7.7)
in a $14 \times 12 \times 5$ matrix
where, $a = 5,6,\ldots,18$
$g = 1,2,\ldots,12$
$i = 1,2,\ldots,5.$

The matrix of transition probabilities reflect the potential movements of students within the system. It is the probability of transition from a grade-state in one year, y, to a grade-state the following year, y+1. Thus a student a years old in grade g for year y will move to grade g' in year y+1. Each i pertains to the risk the student encounters; failing, passing, skipping, graduation and dropping out.

$i = 1$ - repeat grade g (failure), i.e., the movement is from $(a,g) \rightarrow (a+1,g)$.

$i = 2$ - regular promotion (pass) to grade g+1 at age a+1, i.e., the student movement is from $(a,g) \rightarrow (a+1,g+1)$.

$i = 3$ - double promotion to grade g+2 at age a+1, i.e., the student movement is from $(a,g) \rightarrow (a+1,g+2)$.

$i = 4$ - graduation at age a+1, i.e., leaving a transient state (a grade) and entering an absorption state (graduation), $(a,g) \rightarrow (a+1,G)$.

$i = 5$ - dropping out at age a and before age $a+1$ without completing grade g, i.e., the student movement is from $(a,g) \rightarrow (a+1,D)$.

This matrix has a large effect on the projected enrollment tables because of the multiplicative effect of the transition probabilities. No account is taken of possible changes in probabilities from year to year.

Returning Probabilities

$$\underline{R} = \{R(g,a)\} \text{ - returning dropout probabilities, a } 12 \times 14 \text{ matrix} \qquad (7.8)$$

where, $g = 1,2,\ldots,12$

$a = 5,6,\ldots,18.$

$R(g,a)$ shows the returning probability of a student a years old who has dropped out of grade g. This assumes the student returns to the same grade if he re-enters the system.

7.3.4 Endogenous Variables

These output variables depend on the interactions of exogenous and status variables depending on the operation of the system.

Graduate Table

$$\underline{G}_y = \{G_y(a)\} \text{ - graduation table, a } 14 \times 1 \text{ vector} \qquad (7.9)$$

where, $a = 5,6,\ldots,18^+.$

The graduation table gives the number of students who have passed the final grade-state. The bulk of the graduations will occur at ages 16, 17, and 18^{+}. A practical assumption would be the elimination of all the earlier ages. A graduation table will be produced for each projected year.

Permanent Dropout Table

$$\underline{PD}_{y} = \{PD_{y}(g)\} \text{ - permanent dropout table, a } 12 \times 1 \text{ vector} \qquad (7.10)$$

where, $g = 1,2,\ldots,12.$

The permanent dropout table gives the number of dropouts with last grade g who are more than 18 years of age. These are students that we assume enter "adult education" or who leave the formal school system.

7.3.5 Inputs

(1) Age-grade table for the initial year Y to provide a starting point for the model.

(2) Number of years to be simulated with Y+s being the target year.

(3) Parameter matrices $\underline{T}$ and $\underline{R}$.

(4) Status tables $\underline{S}_{y}$, $\underline{D}_{y}$, $\underline{W}_{y}$ for the starting year Y.

(5) Number of births in the year, the entering probability vector $\underline{E}_{y}$ and net migration vector $\underline{I}_{y}$ are each entered for every year to be simulated, i.e., y = Y, Y+1, Y+2, ..., Y+s.

7.3.6 Outputs

(1) Status tables for every simulated year, i.e., $\underline{S}_y$, $\underline{W}_y$, and $\underline{D}_y$, for $y = Y, Y+1, Y+2, \ldots, Y+s$.

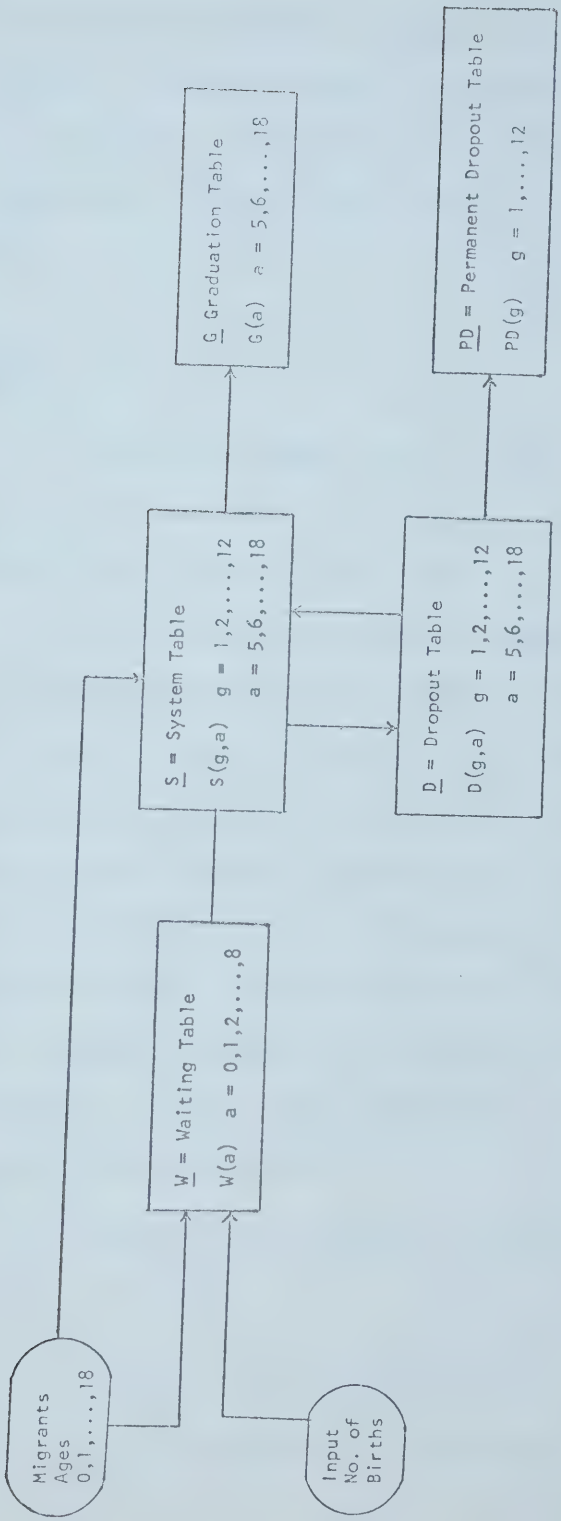
(2) Output tables for every simulated year, i.e., $\underline{G}_y$ and $\underline{PD}_y$ for $y = Y, Y+1, Y+2, \ldots, Y+s$.

Figure 7.1 shows the relationship between many of the processes. Migrants and input number of births enter the pre-school table where they wait to enter the school system. When the time comes the subjects enter the school enrollment table which represents our "school system". All other tables are on the periphery of the system. Migrants can also enter the school enrollment table directly. The student can leave the school enrollment table and enter the dropout table. From the dropout table the student can return to the system or become a permanent dropout. The student may also enter the graduation table from the school enrollment table.

When a student arrives at the entering age he is exposed to only two "risks"; entering the system or waiting. A vector of entering probabilities gives the probability of these two occurrences specific to age. Once the student is in the system he is exposed to another setof risks consisting of passing, failing, skipping, graduating, and dropping out. The risks of passing, failing, or skipping continue until the student graduates or drops out of the system. Each of the risks has its own probability of occurence.

FIGURE 7.1

Flow of Students Into and Out of the School System



The model includes only those students that eventually enter the system, and considers a student who graduates or becomes a final dropout, to be in a state of absorption and ineligible to return to the system. The parameters determining the probabilities of the occurrence of events for any student are assumed constant over time.

7.4 METHODOLOGY

Our simulation assumes that the status tables for any year after the starting year Y, i.e., Y+k, are dependent only on the status table of the previous year, Y+k-1, as well as transition probabilities, returning probabilities of dropouts, and school entering probabilities of pre-school children. Each subject in the cell $\underline{S}_y(g,a)$ in year y transfers to a cell of $\underline{S}_{y+1}$, or $\underline{D}_{y+1}$, or $\underline{G}_{y+1}$ in year y+1. This transfer is independent of the subject's past history and indicates that the student remains in school, drops out or graduates, independent of the pattern he has previously followed. The only condition is that he is in grade = g, age = a, in year = y. Therefore the transition probabilities for this age-grade combination, T(a,g,j), j = 1,2,3,4,5, are conditional

If we denote the five mutually exclusive events by a random variable J with,

J = 1 event of repeating grade g

J = 2 event of promotion to grade g+1

$J = 3$ event of promotion to grade $g+2$

$J = 4$ event of graduation

$J = 5$ event of dropping out,

then,

$$P_J(J = j \text{ in year } y+1 \mid \text{grade} = g, \text{ age} = a, \text{ in year } y) = T(a,g,j) \quad . \tag{7.11}$$

Therefore, the expected number of students for each event j will be

$$S_y(a,g) \times T(a,g,j), \quad \text{where} \quad j = 1,2,3,4,5 \ .$$

Since uncertainties are introduced by probabilistic statements the exact number of students cannot be determined.

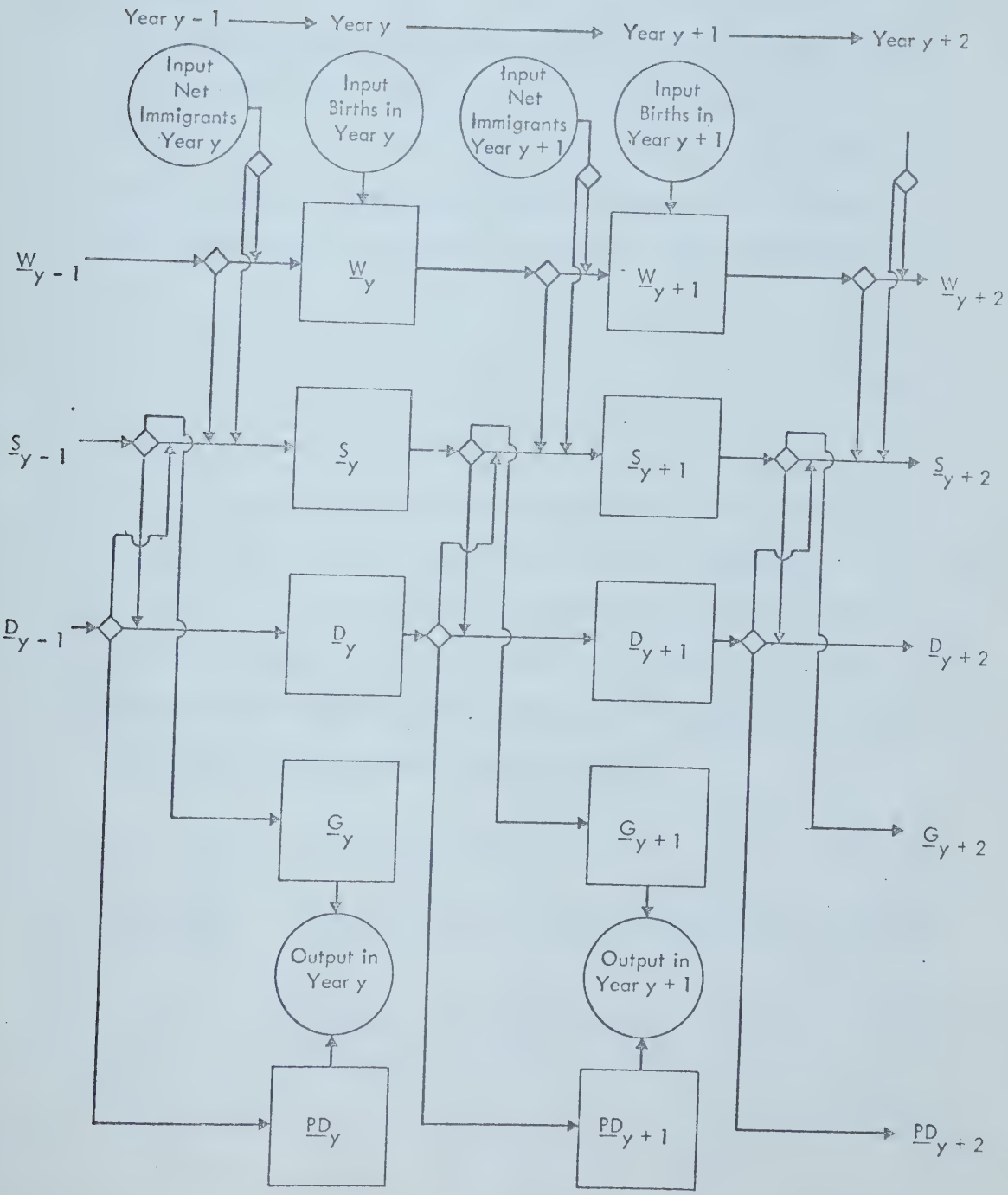
From the foregoing it follows that a Markov process exists as the probabilities for entering into the next state are completely determined by the present state of the system,

$$P(\underline{S}_{y+k} = \underline{F}_{y+k} \mid \underline{S}_{y+k-1} = \underline{F}_{y+k-1}, \ \ldots, \ \underline{S}_y = \underline{F}_y) = P(\underline{S}_{y+k} = \underline{F}_{y+k} \mid \underline{S}_{y+k-1} = \underline{F}_{y+k-1}) \quad . \tag{7.120}$$

Students in any cell of the status tables, in any particular year transfer to the new states in other tables as determined by the random walk model, i.e., only the probabilities specified by the input parameters. The model assumes that the whole system can be described by the three status tables and two output tables with transition between these tables occurring only once a year. The flow of students among the tables from year y to year $y+1$ is illustrated in Figure 7.2.

FIGURE 7.2

Inter Status Tables Flow



7.4.1 Transition Rules

The rules for the movement of a student in table $\underline{S}_y$ in year y to the status and output tables in year y+1 are as follows:

(1) For every student found in the school enrollment table $\underline{S}_y$, a random number from the uniform distribution over the interval (0,1) is generated by the computer and then assigned to the student. A continuous random variable with a uniform distribution has the following probability distribution function;

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & x \leq 0, \quad x \geq 1 \end{cases} \tag{7.13}$$

Based on his assigned random number each student enters the appropriate cell of one of the following tables; $\underline{S}_{y+1}$, $\underline{D}_{y+1}$, or $\underline{G}_{y+1}$. The random numbers are generated by the IBM Scientific Subroutine Package subroutine RANDU. The subroutine is specific to the IBM 360 system and will produce 2^{29} random numbers before repeating. Let the number thus generated be,

$$p \quad (0 < p < 1) \quad ,$$

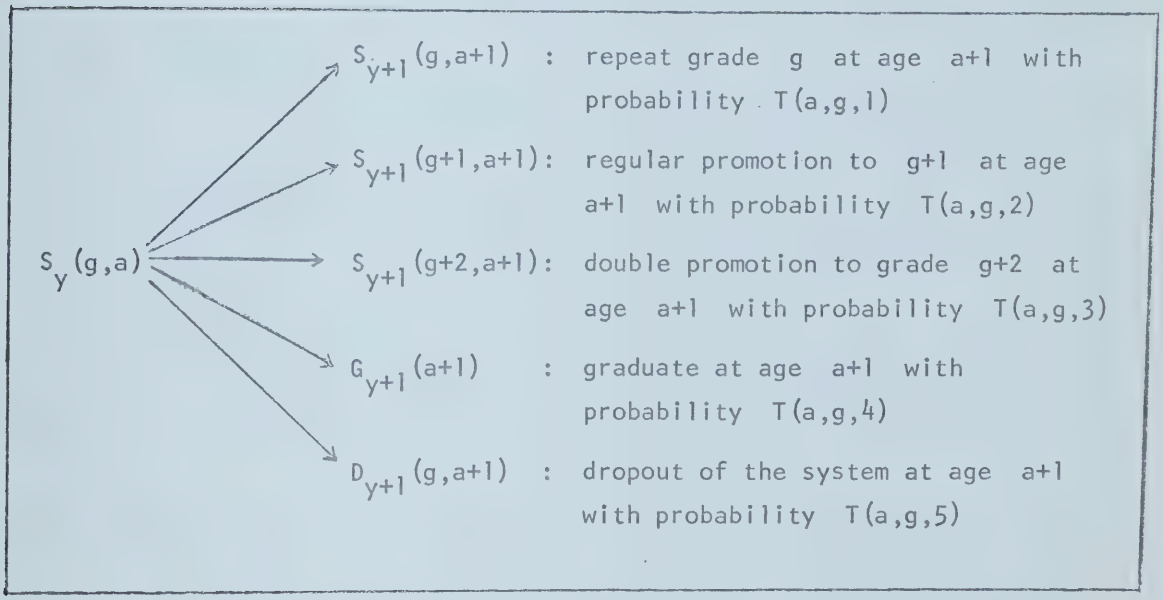
then choose i such that the following relationship is satisfied,

$$\sum_{j=1}^{i-1} T(a,g,j) < p \leq \sum_{j=i+1}^{5} T(a,g,j) \quad . \tag{7.14}$$

Based on this i, the subject transfers to an appropriate cell of $\underline{S}_{y+1}$ school enrollment table, $\underline{G}_{y+1}$, graduation table, $\underline{D}_{y+1}$, dropout table.

We assume that only five transitions are possible by any age and grade:

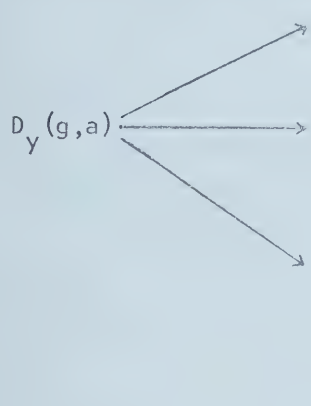
FIGURE 7.3


$S_y(g,a)$
$S_{y+1}(g,a+1)$: repeat grade g at age a+1 with probability T(a,g,1)
$S_{y+1}(g+1,a+1)$: regular promotion to g+1 at age a+1 with probability T(a,g,2)
$S_{y+1}(g+2,a+1)$: double promotion to grade g+2 at age a+1 with probability T(a,g,3)
$G_{y+1}(a+1)$: graduate at age a+1 with probability T(a,g,4)
$D_{y+1}(g,a+1)$: dropout of the system at age a+1 with probability T(a,g,5)

(2) For the subjects already in the dropout table, $\underline{D}_y$ the only transitions possible are to re-enter the school system, stay as a dropout, or permanently dropout due to a movement out of the school age range assumed; i.e., transition is possible only to $\underline{S}_{y+1}$, $\underline{D}_{y+1}$, and $\underline{PD}_{y+1}$. The transition back to the school system $\underline{S}_{y+1}$, is determined by the returning probability given by $\underline{R}$. For each subject in $\underline{D}_y$ a random number p (again generated by RANDU)

is selected and if p is less than $R(g,a)$ he will be transferred to $S_{y+1}(g,a+1)$. He will be transferred to $D_{y+1}(g,a+1)$ if p is greater or equal to $R(g,a)$ and $a+1$ is less than or equal to 18. If $a+1$ is greater than 18 and the subject does not return to $\underline{S}_{y+1}$, he will become a permanent dropout and enter into $PD_{y+1}(g)$.

FIGURE 7.4



$S_{y+1}(g,a+1)$: return to the school system with probability $R(g,a)$

$D_{y+1}(g,a+1)$: stay out of school with probability $[1-R(g,a)]$ if $a+1 \leq 18$

$PD_{y+1}(g)$: become a permanent dropout with probability $[1-R(g,a)]$ if $a+1 > 18$

(3) the subjects in the waiting or pre-school table enter a new waiting table $\underline{W}_{y+1}$ based on the life table $\underline{L}$. Our starting point for this description, $W_{y+1}(0)$, is determined by the number of births in year y.

$$W_y(a) \rightarrow W_{y+1}(a+1) \text{ with probability } L(a) .$$

A portion of $\underline{W}_y(a)$, with $a = 5,6,7,8$, moves to $\underline{S}_{y+1}(1,a+1)$ with the entering probability, $E_{y+1}(a)$. Those aged

over 8 years and still not in $\underline{S}_{y+1}$ will be trans erred to a cell of the dropout table, $D_{y+1}(1,a+1)$, and are treated as though they were dropouts (with last grade equal to one) although they never entered the school system.

(4) The graduation and permanent dropout tables are assumed to be absorption states, once the student enters these states, can never return to the school system.

7.4.2 Identities and Operating Characteristics

The following exact or approximate identities may be observed:

(1) the entries in the cell of the waiting table in year y+1, $(\underline{W}_{y+1})$ are determined by the following components:

(a) The number of live births in year y for W(0).

(b) The number of new entries, $W_y(a)$, times the surviving probabilities of $a \rightarrow a+1$ transitions, i.e., $L(a)$.

(c) The net number of immigrants (a+1) years old in year y+1.

(d) The number of students transferred to the enrollment table $\underline{S}_{y+1}$ which is determined by the entering probabilities, $E_{y+1}(a+1)$.

From these components we get:

$$W_{y+1}(0) = (\text{number of births in } y) + I_{y+1}(0) \qquad (7.15)$$

$$W_{y+1}(a+1) = W_y(a) \times L(a) + I_{y+1}(a+1) , \qquad (7.16)$$

$$a = 0,1,2,3$$

$$W_{y+1}(a+1) = W_y(a) \times L(a)[1.0 - E_{y+1}(a+1)] + I_{y+1}(a+1), \quad a = 4,5 \qquad (7.17)$$

$$W_{y+1}(a+1) = W_y(a) \times L(a) \quad [1.0 - E_{y+1}(a+1)] , \quad a = 6,7 . \qquad (7.18)$$

(2) Entries into any one cell of the enrollment table $\underline{S}_{y+1}$, e.g., $S_{y+1}(g,a+1)$, the number of students in grade g and age $a+1$ in year $y+1$, are determined approximately by the following components:

(a) the expected number of students who repeat grade g which is equal to $S_y(g,a)$ times the failure rate $T(a,g,1)$;

(b) the expected number of passing students is equal to $S_y(g-1,a)$ times the normal progress rate $T(a,g-1,2)$;

(c) the expected number of accelerated (double promotion) students which is equal to $S_y(g-2,a)$ times the rate of skipping grade $g-1$, $T(a,g-2,3)$;

(d) the expected number of students returning from the dropout table $\underline{D}_y$, which is equal to $D_y(g,a)$ times $R(g,a)$;

(e) the number of new students which is equal to $W_{y+1}(a+1)$ times $E_{y+1}(a+1)$ (applicable only for $g = 1$);

(f) an assumption is made that on the average immigrants entering the system are in a grade one lower than the expected grade for the student. For example we would usually expect an 8 year old to be in grade 3, however, an 8 year old new immigrant is assumed to enter grade 2.

Relating all the above components we get:

$$\begin{aligned} S_{y+1}(g,a+1) = {} & W_y(g,a) \times T(a,g,1) + W_y(g-1,a) \times T(a,g-1,2) \\ & + S_y(a,g-2) \times T(a,g-2,3) \\ & + D_y(g,a) \times R(g,a) \\ & + W_{y+1}(a+1) \times E_{y+1}(a+1) + I_{y+1}(a-6) \quad . \end{aligned} \tag{7.19}$$

(3) Entries into the dropout table $\underline{D}_{y+1}$ with last grade g and age a+1, i.e., $D_{y+1}(g,a+1)$ are determined by the following components:

(a) the expected number of students in $D_y(g,a)$ minus the number of subjects who return to the system which equals

$$D_y(g,a) \times [1.0 - R(g,a)] \quad ;$$

(b) the expected number of new dropouts with last grade g and a years old which is equal to $S_y(g,a)$ times the dropout rate $T(a,g,5)$.

Therefore, we get,

$$D_{y+1}(g,a+1) = D_y(g,a) \times [1.0 - R(g,a)] + S_y(g,a) \times T(a,g,5) \quad . \qquad (7.20)$$

(4) The entry to any one cell of the graduate table G_{y+1} is determined by the following components:

(a) the expected number of regular students who graduate from grade 12 which is equal to $S_y(12,a)$ times the graduation rate $T(a,12,4)$;

(b) the expected number of students who graduate from grade 11 which is equal to $S_y(11,a)$ times the graduation rate in grade 11, $T(11,2,4)$.

Therefore we get,

$$G_{y+1}(a+1) = S_y(12,a) \times T(a,12,4) + S_y(11,a) \times T(a,11,4) \quad . \qquad (7.21)$$

(5) The entries to the permanent dropout table $\underline{PD}_{y+1}$ with the last grade g are decided approximately by the number of students in the dropout table with last grade g and age 17 minus those who returned to the system at age 18 which is equal to;

$$D_y(g,17) \times [1.0 - R(g,18)] \quad .$$

Therefore we get,

$$PD_{y+1}(g) = D_y(g,17) \times [1.0 - R(g,17)] \quad . \qquad (7.22)$$

7.4.3 Summary of Inputs

The system requires many inputs as indicated in an earlier section.

(1) The life table uses the following data to determine the survival rate at the different ages:

Age	0	1	2	3	4	5	6	7	8
	.9702	.9988	.9988	.9988	.9995	.9995	.9995	.9995	.9995

(2) All returning probabilities of dropouts are set equal to 0.0 except for those students who are less than 16 years of age and where their last grade is within $\pm$ 1 of the grade range of normal students. For these students the returning probability is set at 0.3.

(3) As the initial starting (or basic) year data we use the student enrollment data of 1969-70. This means that our system table is based on the enrollment table of 1969-70. The waiting and dropout tables used are based on the 1950 to 1970 simulation which was used for parameter estimation.

(4) Yearly external parameter data (Figure 7.5) consisting of the number of births and school entering rates for ages 5 to 8 have been calculated and projected. The projected number of births have been calculated separately for females and males. The number of births double from 1970-71 to 2004-05.

FIGURE 7.5

Yearly Input Data

Year	Number of Births*		Entering Rate			
	Female	Male	5	6	7	8
1970-1971	163	169	0.3700	0.9628	0.2200	0.0200
1971-1972	175	181	0.4000	0.9593	0.2200	0.0200
1972-1973	180	186	0.4200	0.9573	0.2200	0.0200
1973-1974	186	192	0.4300	0.9558	0.2200	0.0200
1974-1975	191	198	0.4400	0.955	0.2200	0.0200
1975-1976	197	203	0.4500	0.9542	0.2200	0.0200
1976-1977	202	209	0.4600	0.9534	0.2200	0.0200
1977-1978	209	215	0.4700	0.9525	0.2200	0.0200
1978-1979	215	222	0.4800	0.9516	0.2200	0.0200
1979-1980	221	228	0.4900	0.9507	0.2200	0.0200
1980-1981	227	235	0.5000	0.9497	0.2200	0.0200
1981-1982	234	241	0.5100	0.9487	0.2200	0.0200
1982-1983	238	245	0.5200	0.9477	0.2200	0.0200
1983-1984	242	250	0.5300	0.9466	0.2200	0.0200
1984-1985	246	254	0.5400	0.9454	0.2200	0.0200
1985-1986	250	258	0.5500	0.9443	0.2200	0.0200
1986-1987	254	262	0.5600	0.9430	0.2200	0.0200
1987-1988	258	266	0.5700	0.9417	0.2200	0.0200
1988-1989	262	270	0.5800	0.9404	0.2200	0.0200
1989-1990	266	274	0.5900	0.9386	0.2200	0.0200
1990-1991	270	278	0.6500	0.9375	0.2200	0.0200
1991-1992	274	283	0.6100	0.9359	0.2200	0.0200
1992-1993	278	287	0.6200	0.9342	0.2200	0.0200
1993-1994	282	291	0.6300	0.9325	0.2200	0.0200
1994-1995	286	295	0.6400	0.9307	0.2200	0.0200
1995-1996	290	299	0.6500	0.9288	0.2200	0.0200
1996-1997	294	303	0.6600	0.9267	0.2200	0.0200
1997-1998	298	307	0.6700	0.9246	0.2200	0.0200
1998-1999	302	311	0.6800	0.8223	0.2200	0.0200
1999-2000	306	315	0.6900	0.9199	0.2200	0.0200
2000-2001	310	319	0.7000	0.9173	0.2200	0.0200
2001-2002	314	323	0.7100	0.9145	0.2200	0.0200
2002-2003	318	327	0.7200	0.9116	0.2200	0.0200
2003-2004	322	331	0.7300	0.9084	0.2200	0.0200
2004-2005	326	335	0.7400	0.9050	0.2200	0.0200

* Unit = 100, based on calendar year.

The entrance rates are the proportion of students at a specified age who enter school in that year. For 1970-71 we have 0.37 (37%) of the five year olds entering school, 0.9628 of the six year olds entering, 0.22 of the seven year olds entering, and 0.02 of the eight year olds. We project that the rate of five year olds entering school will increase from 0.37 in 1970-71 to 0.74 in 2004-05, based on the trend that sees the lowering of school age requirements. The impact of compulsory kindergartens would probably see an increase in this proportion. The entering rates for seven and eight year olds is constant.

It is assumed that the net migration effect is zero.

Figures 7.6 and 7.7 show the transition parameters for males and females respectively. These are essentially the same parameters as used in Chapter 6, Markov Chain Model, in order that a comparison of the two models may be made except that some modifications are made for lower and higher grades to fit the real data between 1950-51 and 1969-70. For illustration purposes we will refer to Figure 7.6. For all "appropriate" age and grade combinations the transition parameters for failure (repeat), pass (normal), skip, graduate, and dropout have been determined. For example, students age eleven in grade 6 have a 0.03 probability of failure, a 0.96 probability of passing, nil probabilities of skipping or graduating and a 0.01 probability of being a dropout. Across each row the summation of all probabilities is 1.0. We found that the number of "skippers" is so small that we assume it is always zero. We assume that it is impossible to graduate until

FIGURE 7.6

Transition Parameter Set 1, Male

Age	Grade	Repeat	Normal	Skip	Graduate	Dropout
5	1	0.07500	0.92000	0.0	0.0	0.00500
5	2	0.05500	0.94000	0.0	0.0	0.00500
5	3	0.04000	0.95000	0.0	0.0	0.01000
6	1	0.07500	0.92000	0.0	0.0	0.00500
6	2	0.05500	0.94000	0.0	0.0	0.00500
6	3	0.04000	0.95000	0.0	0.0	0.01000
7	1	0.07500	0.92000	0.0	0.0	0.00500
7	2	0.05500	0.94000	0.0	0.0	0.00500
7	3	0.04000	0.95000	0.0	0.0	0.01000
7	4	0.04000	0.95000	0.0	0.0	0.01000
8	1	0.07500	0.92000	0.0	0.0	0.00500
8	2	0.05500	0.94000	0.0	0.0	0.00500
8	3	0.04000	0.95000	0.0	0.0	0.01000
8	4	0.04000	0.95000	0.0	0.0	0.01000
8	5	0.04000	0.95000	0.0	0.0	0.01000
9	2	0.05500	0.94000	0.0	0.0	0.00500
9	3	0.04000	0.95000	0.0	0.0	0.01000
9	4	0.04000	0.95000	0.0	0.0	0.01000
9	5	0.04000	0.95000	0.0	0.0	0.01000
9	6	0.03000	0.96000	0.0	0.0	0.01000
10	3	0.04000	0.95000	0.0	0.0	0.01000
10	4	0.04000	0.95000	0.0	0.0	0.01000
10	5	0.04000	0.95000	0.0	0.0	0.01000
10	6	0.03000	0.96000	0.0	0.0	0.01000
10	7	0.05500	0.92000	0.0	0.0	0.02500
11	4	0.04000	0.95000	0.0	0.0	0.01000
11	5	0.04000	0.95000	0.0	0.0	0.01000
11	6	0.03000	0.96000	0.0	0.0	0.01000
11	7	0.05500	0.92000	0.0	0.0	0.02500
11	8	0.04000	0.92000	0.0	0.0	0.04000
12	5	0.04000	0.95000	0.0	0.0	0.01000
12	6	0.03000	0.96000	0.0	0.0	0.01000
12	7	0.05500	0.92000	0.0	0.0	0.02500
12	8	0.04000	0.92000	0.0	0.0	0.04000
12	9	0.05500	0.80000	0.0	0.0	0.14500
13	6	0.03000	0.96000	0.0	0.0	0.01000
13	7	0.05500	0.92000	0.0	0.0	0.02500
13	8	0.04000	0.92000	0.0	0.0	0.04000
13	9	0.05500	0.80000	0.0	0.0	0.14500
13	10	0.06000	0.85000	0.0	0.0	0.09000
14	7	0.05500	0.92000	0.0	0.0	0.02500
14	8	0.04000	0.92000	0.0	0.0	0.04000
14	9	0.05500	0.80000	0.0	0.0	0.14500
14	10	0.06000	0.85000	0.0	0.0	0.09000
14	11	0.09000	0.81000	0.0	0.0	0.10000
15	8	0.04000	0.92000	0.0	0.0	0.04000
15	9	0.05500	0.80000	0.0	0.0	0.14500
15	10	0.06000	0.85000	0.0	0.0	0.09000
15	11	0.09000	0.81000	0.0	0.0	0.10000
15	12	0.29000	0.00000	0.0	0.6000	0.11000
16	9	0.05500	0.80000	0.0	0.0	0.14500
16	10	0.06000	0.85000	0.0	0.0	0.09000
16	11	0.09000	0.81000	0.0	0.0	0.10000
16	12	0.05000	0.00000	0.0	0.85000	0.10000
17	10	0.06000	0.85000	0.0	0.0	0.09000
17	11	0.09000	0.81000	0.0	0.0	0.10000
17	12	0.05000	0.00000	0.0	0.85000	0.10000
18	10	0.05000	0.90000	0.0	0.0	0.05000
18	11	0.05000	0.90000	0.0	0.0	0.05000
18	12	0.05000	0.00000	0.0	0.6000	0.35000

Note: All other probabilities are set to 0.0 except for dropout probabilities which are equal to 1.0.

FIGURE 7.7

Transition Parameter Set 2, Female

Age	Grade	Repeat	Normal	Skip	Graduate	Dropout
5	1	0.01500	0.98500	0.0	0.0	0.0
5	2	0.03500	0.96000	0.0	0.0	0.00500
5	3	0.02500	0.97000	0.0	0.0	0.00500
6	1	0.04500	0.95000	0.0	0.0	0.00500
6	2	0.03500	0.96000	0.0	0.0	0.00500
6	3	0.02500	0.97000	0.0	0.0	0.00500
7	1	0.04500	0.95000	0.0	0.0	0.00500
7	2	0.03500	0.96000	0.0	0.0	0.00500
7	3	0.02500	0.97000	0.0	0.0	0.00500
7	4	0.03000	0.96000	0.0	0.0	0.01000
8	1	0.04500	0.95000	0.0	0.0	0.00500
8	2	0.03500	0.96000	0.0	0.0	0.00500
8	3	0.02500	0.97000	0.0	0.0	0.00500
8	4	0.03000	0.96000	0.0	0.0	0.01000
8	5	0.03000	0.96000	0.0	0.0	0.01000
9	2	0.03500	0.96000	0.0	0.0	0.00500
9	3	0.02500	0.97000	0.0	0.0	0.00500
9	4	0.03000	0.96000	0.0	0.0	0.01000
9	5	0.03000	0.96000	0.0	0.0	0.01000
9	6	0.03500	0.95000	0.0	0.0	0.01500
10	3	0.02500	0.97000	0.0	0.0	0.00500
10	4	0.03000	0.96000	0.0	0.0	0.01000
10	5	0.03000	0.96000	0.0	0.0	0.01000
10	6	0.03500	0.95000	0.0	0.0	0.01500
10	7	0.05500	0.93000	0.0	0.0	0.01500
11	4	0.03000	0.96000	0.0	0.0	0.01000
11	5	0.03000	0.96000	0.0	0.0	0.01000
11	6	0.03500	0.95000	0.0	0.0	0.01500
11	7	0.05500	0.93000	0.0	0.0	0.01500
11	8	0.05000	0.93000	0.0	0.0	0.02000
12	5	0.03000	0.96000	0.0	0.0	0.01000
12	6	0.03500	0.95000	0.0	0.0	0.01500
12	7	0.05500	0.93000	0.0	0.0	0.01500
12	8	0.05000	0.93000	0.0	0.0	0.02000
12	9	0.08000	0.82000	0.0	0.0	0.10000
13	6	0.03500	0.95000	0.0	0.0	0.01500
13	7	0.03500	0.95000	0.0	0.0	0.01500
13	8	0.05000	0.93000	0.0	0.0	0.02000
13	9	0.08000	0.82000	0.0	0.0	0.10000
13	10	0.06000	0.84000	0.0	0.0	0.10000
14	7	0.03000	0.95000	0.0	0.0	0.02000
14	8	0.03500	0.95000	0.0	0.0	0.01500
14	9	0.02000	0.95000	0.0300	0.0	0.00000
14	10	0.05000	0.80000	0.0	0.0	0.15000
14	11	0.09000	0.75000	0.0	0.0	0.16000
15	8	0.03000	0.95000	0.0	0.0	0.02000
15	9	0.03000	0.95000	0.0	0.0	0.02000
15	10	0.02000	0.95000	0.0	0.0	0.03000
15	11	0.02000	0.95000	0.0	0.0	0.03000
15	12	0.05000	0.00000	0.0	0.9000	0.05000
16	9	0.03000	0.95000	0.0	0.0	0.02000
16	10	0.02000	0.95000	0.0	0.0	0.03000
16	11	0.04000	0.90000	0.0	0.0	0.06000
16	12	0.05000	0.00000	0.0	0.9000	0.05000
17	10	0.05000	0.40000	0.0	0.0	0.55000
17	11	0.05000	0.35000	0.0	0.0	0.60000
17	12	0.02000	0.00000	0.0	0.95000	0.03000
18	10	0.05000	0.30000	0.0	0.0	0.65000
18	11	0.05000	0.30000	0.0	0.0	0.65000
18	12	0.05000	0.00000	0.0	0.9000	0.05000

Note: All other probabilities are set to 0.0 except for droput probabilities which are equal to 1.0.

a student reaches grade 12. As a consequence our graduation probabilities are all zero until the age of 15. Although in theory a student is prevented by law from dropping out of school before age 16, we tend to use our dropout category as a "catchall" that includes invalids, real dropouts, mortalities, etc.

Figures 7.8 and 7.9 show another set of transition parameters for males and females respectively. These parameters have been developed from the method of successive subtractions referred to in Appendix B, with the same modification as in the previous paragraph. For example, Figure 7.9, students age 14 in grade 11 show a nil probability of failing and a .99999 probability of passing. This is probably accounted for by few students fitting into this age-grade cell and each of these performing well in school. This method accounts for skippers for each grade, giving more detailed transition parameters then the previous method.

7.5 PROJECTIONS OF ENROLLMENT

Starting with 1971, projections of enrollment were made for every year up to the school year 2004-05 for age and grade specific series under two different sets of parameters.

(1) Figures 7.10 to 7.13 give projections with the same parameters used in Chapter 6. Figure 7.10 gives the projections of school enrollment by age for males. There are some particularly interesting results found in the Total column. For 1970-71 the total enrollment is projected to be 209,200 dropping off continuously (with a slight jump in 1975-76) until it reaches the

FIGURE 7.8

Transition Parameter Set 3, Male

Age	Grade	Repeat	Normal	Skip	Graduate	Dropout
5	1	0.10100	0.89900	0.0	0.0	0.0
5	2	0.08700	0.91000	0.00300	0.0	0.0
5	3	0.07400	0.92500	0.00100	0.0	0.0
6	1	0.10100	0.89900	0.0	0.0	0.0
6	2	0.08700	0.91000	0.00300	0.0	0.0
6	3	0.07400	0.92500	0.00100	0.0	0.0
7	1	0.10100	0.89900	0.0	0.0	0.0
7	2	0.08700	0.91000	0.00300	0.0	0.0
7	3	0.07400	0.92500	0.00100	0.0	0.0
7	4	0.07900	0.81900	0.00200	0.0	0.0
8	1	0.10100	0.89900	0.0	0.0	0.0
8	2	0.08700	0.9100	0.00300	0.0	0.0
8	3	0.07400	0.92500	0.00100	0.0	0.0
8	4	0.07900	0.91900	0.00200	0.0	0.0
8	5	0.07300	0.92700	0.0	0.0	0.0
9	2	0.08700	0.91000	0.00300	0.0	0.0
9	3	0.07400	0.92500	0.00100	0.0	0.0
9	4	0.07900	0.91900	0.00200	0.0	0.0
9	5	0.07300	0.92700	0.0	0.0	0.0
9	6	0.07300	0.92700	0.0	0.0	0.0
10	3	0.07400	0.92500	0.00100	0.0	0.0
10	4	0.07900	0.91900	0.00200	0.0	0.0
10	5	0.07300	0.92700	0.0	0.0	0.0
10	6	0.07300	0.92700	0.0	0.0	0.0
10	7	0.09500	0.90500	0.0	0.0	0.0
11	4	0.07900	0.91900	0.00200	0.0	0.0
11	5	0.07300	0.92700	0.0	0.0	0.0
11	6	0.07300	0.92700	0.0	0.0	0.0
11	7	0.09500	0.90500	0.0	0.0	0.0
11	8	0.07500	0.92500	0.0	0.0	0.0
12	5	0.07300	0.92700	0.0	0.0	0.0
12	6	0.07300	0.92700	0.0	0.0	0.0
12	7	0.09500	0.90500	0.0	0.0	0.0
12	8	0.07500	0.92500	0.0	0.0	0.0
12	9	0.07000	0.92200	0.00800	0.0	0.0
13	6	0.07300	0.92700	0.0	0.0	0.0
13	7	0.09500	0.90500	0.0	0.0	0.0
13	8	0.07500	0.92500	0.0	0.0	0.0
13	9	0.07000	0.92200	0.00800	0.0	0.0
13	10	0.00700	0.92400	0.06900	0.0	0.0
14	7	0.09500	0.90500	0.0	0.0	0.0
14	8	0.07500	0.92500	0.0	0.0	0.0
14	9	0.07000	0.92200	0.00800	0.0	0.0
14	10	0.00700	0.92400	0.06900	0.0	0.0
14	11	0.00000	0.99999	0.0	0.0	0.00001
15	8	0.07500	0.92500	0.0	0.0	0.0
15	9	0.07000	0.92200	0.00800	0.0	0.0
15	10	0.00700	0.92400	0.06900	0.0	0.0
15	11	0.00000	0.99999	0.0	0.0	0.00001
15	12	0.00000	0.00000	0.0	0.66500	0.33500
16	9	0.07000	0.82200	0.00800	0.0	0.10000
16	10	0.00700	0.82400	0.06900	0.0	0.10000
16	11	0.00000	0.99999	0.0	0.0	0.00001
16	12	0.20000	0.00000	0.0	0.66500	0.13500
17	10	0.00700	0.92400	0.06900	0.0	0.0
17	11	0.00000	0.90000	0.0	0.0	0.10000
17	12	0.00000	0.00000	0.0	0.60000	0.40000
18	10	0.00700	0.92400	0.06900	0.0	0.0
18	11	0.00000	0.99999	0.0	0.0	0.00001
18	12	0.00000	0.00000	0.0	0.66500	0.33500

Note: All other probabilities are set to 0.0 except for dropout probabilities which are equal to 1.0.

FIGURE 7.9

Transition Parameter Set 4, Female

Age	Grade	Repeat	Normal	Skip	Graduate	Dropout
5	1	0.01500	0.98500	0.0	0.0	0.0
5	2	0.03700	0.96000	0.00300	0.0	0.0
5	3	0.02500	0.97400	0.00100	0.0	0.0
6	1	0.04500	0.95500	0.0	0.0	0.0
6	2	0.03700	0.96000	0.00300	0.0	0.0
6	3	0.02500	0.07400	0.00100	0.0	0.0
7	1	0.04500	0.95500	0.0	0.0	0.0
7	2	0.03500	0.96200	0.00300	0.0	0.0
7	3	0.02400	0.97500	0.00100	0.0	0.0
7	4	0.02900	0.96900	0.00200	0.0	0.0
8	1	0.10100	0.89900	0.0	0.0	0.0
8	2	0.04700	0.95000	0.00300	0.0	0.0
8	3	0.03400	0.96500	0.00100	0.0	0.0
8	4	0.02900	0.06900	0.00200	0.0	0.0
8	5	0.03300	0.96700	0.0	0.0	0.0
9	2	0.08700	0.91000	0.00300	0.0	0.0
9	3	0.07400	0.92500	0.00100	0.0	0.0
9	4	0.07900	0.91900	0.00200	0.0	0.0
9	5	0.07300	0.92700	0.0	0.0	0.0
9	6	0.07300	0.92700	0.0	0.0	0.0
10	3	0.07400	0.92500	0.00100	0.0	0.0
10	4	0.07900	0.91900	0.00200	0.0	0.0
10	5	0.07300	0.92700	0.0	0.0	0.0
10	6	0.07300	0.92700	0.0	0.0	0.0
10	7	0.09500	0.90500	0.0	0.0	0.0
11	4	0.07900	0.91900	0.00200	0.0	0.0
11	5	0.07300	0.92700	0.0	0.0	0.0
11	6	0.07300	0.92700	0.0	0.0	0.0
11	7	0.09500	0.90500	0.0	0.0	0.0
11	8	0.07500	0.92500	0.0	0.0	0.0
12	5	0.07300	0.92700	0.0	0.0	0.0
12	6	0.07300	0.92700	0.0	0.0	0.0
12	7	0.09500	0.90500	0.0	0.0	0.0
12	8	0.07500	0.92500	0.0	0.0	0.0
12	9	0.07000	0.92200	0.00800	0.0	0.0
13	6	0.07300	0.92700	0.0	0.0	0.0
13	7	0.09500	0.90500	0.0	0.0	0.0
13	8	0.07500	0.92500	0.0	0.0	0.0
13	9	0.07000	0.92200	0.00800	0.0	0.0
13	10	0.00700	0.92400	0.06900	0.0	0.0
14	7	0.09500	0.90500	0.0	0.0	0.0
14	8	0.07500	0.92500	0.0	0.0	0.0
14	9	0.07000	0.92200	0.00800	0.0	0.0
14	10	0.00700	0.92400	0.06900	0.0	0.0
14	11	0.00000	0.99999	0.0	0.0	0.00001
15	8	0.07500	0.92500	0.0	0.0	0.0
15	9	0.07000	0.92200	0.00800	0.0	0.0
15	10	0.00700	0.92400	0.06900	0.0	0.0
15	11	0.00000	0.99999	0.0	0.0	0.00001
15	12	0.00000	0.00000	0.0	0.66500	0.33500
16	9	0.07000	0.82200	0.00800	0.0	0.10000
16	10	0.00700	0.82400	0.06900	0.0	0.10000
16	11	0.00000	0.99999	0.0	0.0	0.00001
16	12	0.20000	0.00000	0.0	0.66500	0.13500
17	10	0.05000	0.40000	0.0	0.0	0.55000
17	11	0.05000	0.35000	0.0	0.0	0.60000
17	12	0.05000	0.00000	0.0	0.91000	0.04000
18	10	0.05000	0.30000	0.0	0.0	0.65000
18	11	0.05000	0.30000	0.0	0.0	0.65000
18	12	0.05000	0.00000	0.0	0.90000	0.05000

Note: All other probabilities are set to 0.0 except for dropout probabilities which are equal to 1.0.

FIGURE 7.10

Projection by Simulation, Age Specific, Male, Parameter Set 1 (Figure 7.6)

(Unit = 100)

Year \ Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970-1971	60	174	183	190	185	188	182	174	170	154	150	142	101	39	2092
1971-1972	60	159	175	183	189	185	189	178	170	166	138	139	104	42	2077
1972-1973	63	147	160	175	182	191	182	187	173	159	149	121	103	49	2041
1973-1974	64	148	147	160	173	182	188	179	184	172	145	136	93	51	2022
1974-1975	66	147	148	146	157	173	177	186	177	178	145	130	94	50	1974
1975-1976	73	147	147	148	144	155	171	173	183	170	168	140	103	59	1981
1976-1977	80	159	147	147	147	143	156	169	172	169	154	156	109	54	1962
1977-1978	84	171	158	144	143	146	143	153	167	163	153	140	125	64	1954
1978-1979	89	175	171	158	143	143	145	141	152	158	150	144	105	73	1947
1979-1980	94	180	176	170	158	144	144	142	135	145	140	142	118	73	1961
1980-1981	98	186	181	174	168	158	144	142	141	131	133	137	110	79	1982
1981-1982	103	189	186	181	174	166	158	142	138	136	117	113	108	81	1992
1982-1983	108	195	186	185	181	174	166	156	138	134	126	110	81	75	2015
1983-1984	113	203	195	186	184	176	169	166	152	129	120	120	83	64	2060
1984-1985	119	209	203	195	184	183	178	165	158	141	116	107	92	62	2112
1985-1986	125	215	210	201	192	183	183	176	160	150	131	106	78	65	2175
1986-1987	131	221	214	204	200	192	182	179	170	155	141	116	77	53	2235
1987-1988	135	227	222	213	204	198	189	181	172	163	144	130	82	47	2307
1988-1989	140	231	226	222	212	205	199	188	171	159	148	134	84	48	2367
1989-1990	145	235	232	224	219	211	204	196	181	166	148	143	96	61	2461
1990-1991	149	238	235	231	223	215	209	200	190	169	156	146	95	64	2520
1991-1992	154	243	237	234	229	215	216	206	199	174	153	152	107	57	2576
1992-1993	162	246	243	234	234	227	214	214	199	176	162	141	108	72	2632
1993-1994	165	249	248	243	230	232	226	214	208	189	161	146	106	66	2683
1994-1995	170	254	251	246	241	227	231	222	203	199	174	149	103	62	2732
1995-1996	175	259	255	250	243	238	224	228	220	192	181	157	112	61	2795
1996-1997	181	263	258	255	251	245	239	222	225	210	177	171	104	66	2867
1997-1998	186	265	263	258	252	248	238	233	217	208	186	158	111	64	2887
1998-1999	191	269	264	260	256	247	246	238	230	204	196	169	112	74	2956
1999-2000	197	274	271	260	259	255	247	241	227	212	192	185	122	76	3018
2000-2001	202	276	274	266	261	254	257	244	236	216	189	180	121	87	3063
2001-2002	208	281	277	273	263	259	249	250	236	221	202	177	155	83	3094
2002-2003	214	286	282	277	272	263	256	249	247	226	210	190	115	77	3164
2003-2004	220	287	286	279	275	273	260	251	244	233	206	201	133	74	3222
2004-2005	226	293	288	284	277	274	271	258	247	231	213	197	133	90	3282

FIGURE 7.11

Projection by Simulation, Age Specific, Female, Parameter Set 2 (Figure 7.7)

(Unit = 100)

Year \ Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970-1971	56	169	176	179	177	180	171	165	160	158	153	144	88	10	1986
1971-1972	58	149	169	175	179	176	177	163	160	154	152	146	100	11	1973
1972-1973	61	142	147	169	175	178	175	176	167	154	151	149	112	11	1967
1973-1974	60	142	141	145	170	175	177	174	175	162	156	145	111	16	1949
1974-1975	64	137	142	139	144	170	175	174	172	173	161	152	114	14	1931
1975-1976	71	143	136	142	141	142	170	170	173	161	169	162	119	12	1911
1976-1977	78	153	142	135	142	141	142	164	170	170	155	162	121	12	1887
1977-1978	82	165	154	142	134	142	138	141	163	160	171	152	130	28	1902
1978-1979	86	170	165	154	142	135	140	133	139	159	158	171	125	18	1895
1979-1980	90	175	170	165	151	142	135	135	133	134	154	154	136	25	1899
1980-1981	95	180	176	170	164	149	142	132	137	129	128	155	126	23	1906
1981-1982	99	186	181	176	169	164	148	138	132	131	123	130	123	14	1914
1982-1983	105	190	184	180	175	168	164	148	138	127	129	124	102	21	1955
1983-1984	110	197	191	186	179	174	168	163	147	134	125	131	92	20	2017
1984-1985	115	203	198	191	185	178	175	164	161	138	127	120	89	19	2063
1985-1986	120	208	202	196	189	184	173	171	166	155	130	131	87	17	2129
1986-1987	127	214	209	203	194	183	182	174	170	155	147	134	97	15	2204
1987-1988	131	221	214	208	203	194	180	178	174	164	150	148	102	16	2283
1988-1989	136	225	222	214	206	203	192	179	177	166	156	151	105	16	2348
1989-1990	141	228	225	221	212	205	200	189	178	168	158	152	105	19	2401
1990-1991	145	232	226	224	220	209	203	198	186	171	159	157	110	25	2465
1991-1992	150	236	231	225	220	218	210	199	193	175	164	158	115	21	2515
1992-1993	157	239	235	230	225	218	216	209	198	189	168	166	109	21	2580
1993-1994	160	243	239	236	230	224	217	214	207	191	177	168	109	24	2639
1994-1995	165	247	242	239	236	225	222	212	214	200	179	178	121	21	2701
1995-1996	170	251	247	239	240	235	225	220	211	204	187	175	117	27	2748
1996-1997	175	255	251	247	240	238	234	222	222	202	192	191	111	28	2808
1997-1998	180	259	256	251	247	236	234	233	220	207	190	193	126	20	2852
1998-1999	186	263	258	256	249	244	237	230	228	204	199	190	136	24	2914
1999-2000	191	266	264	257	256	248	241	230	228	218	195	193	122	25	2934
2000-2001	196	269	265	263	256	253	246	239	228	221	207	192	132	21	2988
2001-2002	202	273	269	264	261	252	252	243	231	215	213	208	130	21	3034
2002-2003	208	277	268	269	265	260	252	251	236	221	205	212	136	22	3082
2003-2004	213	281	276	270	267	262	258	247	251	226	209	202	137	22	3122
2004-2005	219	285	280	276	267	264	262	252	248	227	212	214	131	32	3169

FIGURE 7.12

Projection by Simulation, Grade Specific, Male, Parameter Set 1 (Figure 7.6)

(Unit = 100)

Year \ Grade	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL	GRADU-ATES
1970-1971	195	203	196	190	188	187	185	180	159	143	143	123	2092	114
1971-1972	173	201	195	190	193	192	186	183	172	138	131	123	2077	97
1972-1973	172	161	203	189	187	193	205	176	171	140	126	118	2041	103
1973-1974	156	173	160	203	190	182	188	202	184	150	119	115	2022	101
1974-1975	161	149	172	158	207	179	194	179	196	148	134	97	1974	93
1975-1976	164	161	148	166	158	200	181	185	194	168	134	122	1981	73
1976-1977	173	164	160	155	162	154	202	175	194	152	158	113	1962	91
1977-1978	187	171	158	153	157	158	162	198	173	161	140	136	1954	95
1978-1979	203	171	170	165	150	150	160	159	202	149	143	125	1947	108
1979-1980	197	202	174	163	162	153	155	151	167	166	142	129	1961	88
1980-1981	203	196	193	173	164	163	158	152	155	147	156	122	1982	107
1981-1982	215	191	197	189	173	161	171	146	151	135	131	132	1992	85
1982-1983	226	195	189	204	182	175	163	165	151	130	118	117	2015	94
1983-1984	226	222	184	191	204	176	175	158	166	131	123	104	2060	80
1984-1985	231	219	216	192	184	201	186	157	162	132	121	111	2112	73
1985-1986	238	227	210	221	190	178	207	181	157	136	122	108	2175	87
1986-1987	248	226	225	214	211	184	182	207	176	133	124	105	2235	73
1987-1988	253	236	224	226	212	205	187	177	207	150	118	112	2307	77
1988-1989	261	244	234	223	221	210	212	173	179	170	143	97	2367	82
1989-1990	262	256	237	226	230	211	215	207	177	163	152	125	2461	71
1990-1991	254	271	244	234	227	218	219	209	203	162	149	130	2520	95
1991-1992	274	245	263	246	232	219	217	219	208	171	154	128	2576	94
1992-1993	279	268	242	257	241	227	217	215	225	165	163	133	2632	99
1993-1994	274	271	265	240	253	239	226	219	219	184	151	142	2683	98
1994-1995	284	274	260	263	241	242	241	220	220	185	170	132	2732	113
1995-1996	285	277	276	261	250	238	252	232	217	189	169	149	2795	96
1996-1997	287	280	276	276	261	249	245	243	246	177	183	144	2867	115
1997-1998	291	281	279	271	265	257	247	236	248	203	165	144	2887	116
1998-1999	290	287	282	270	267	261	270	239	243	211	192	144	2956	111
1999-2000	296	290	290	281	259	266	263	258	238	215	193	169	3018	108
2000-2001	306	291	280	293	277	250	279	255	255	212	193	172	3063	126
2001-2002	317	293	284	279	293	270	261	263	253	217	199	165	3094	128
2002-2003	327	296	299	278	279	286	288	251	272	219	208	161	3164	128
2003-2004	324	318	282	299	277	273	301	278	254	237	208	171	3222	128
2004-2005	324	319	308	279	298	273	293	279	296	203	223	187	3282	137

FIGURE 7.13

Projection by Simulation, Grade Specific, Female, Parameter Set 2 (Figure 7.7)

(Unit = 100)

Year \ Grade	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL	GRADU-ATES
1970-1971	182	185	183	187	173	176	168	168	163	154	138	109	1986	118
1971-1972	156	183	187	181	181	174	176	170	156	155	137	117	1973	102
1972-1973	151	157	180	191	178	186	174	165	168	152	146	119	1967	106
1973-1974	148	153	152	177	191	183	187	174	159	159	149	117	1949	113
1974-1975	148	148	150	152	174	193	186	185	171	151	152	121	1931	110
1975-1976	152	151	148	154	145	182	189	187	172	164	141	126	1911	114
1976-1977	166	150	150	147	155	150	179	197	175	157	158	103	1887	118
1977-1978	175	168	149	151	140	155	147	185	188	158	159	127	1902	92
1978-1979	181	172	169	146	156	137	158	142	183	179	148	124	1895	116
1979-1980	182	187	167	169	145	157	138	153	143	167	175	116	1899	115
1980-1981	192	184	183	166	166	150	157	139	151	139	150	129	1906	109
1981-1982	194	196	187	175	171	165	147	154	137	140	133	115	1914	120
1982-1983	202	194	193	188	172	172	172	139	154	130	135	104	1955	110
1983-1984	207	210	187	197	184	174	175	163	142	148	122	108	2017	98
1984-1985	212	211	207	192	194	182	176	170	164	126	133	96	2063	100
1985-1986	218	213	210	204	192	188	184	176	165	156	119	104	2129	91
1986-1987	228	225	205	211	201	187	186	189	174	154	141	103	2204	100
1987-1988	232	237	216	205	208	202	186	184	190	163	147	113	2283	96
1988-1989	235	240	228	220	203	209	202	183	180	177	150	121	2348	102
1989-1990	240	236	242	223	221	207	197	198	181	174	164	118	2401	113
1990-1991	241	242	234	241	225	216	208	198	190	169	166	135	2465	109
1991-1992	248	242	236	236	237	227	221	194	197	177	161	139	2515	126
1992-1993	254	249	240	231	234	238	231	219	197	186	167	134	2580	130
1993-1994	256	253	251	236	233	233	240	229	216	188	163	141	2639	125
1994-1995	257	259	250	255	227	237	232	236	224	209	172	143	2701	132
1995-1996	265	259	251	248	261	229	236	233	232	202	188	144	2748	129
1996-1997	266	265	260	261	244	258	232	228	236	219	188	151	2808	139
1997-1998	272	269	260	265	254	246	255	233	227	219	203	149	2852	141
1998-1999	278	272	266	257	271	252	249	241	236	212	205	175	2914	136
1999-2000	276	287	266	264	261	264	252	253	242	221	190	158	2934	164
2000-2001	287	269	285	267	267	257	268	254	245	232	197	160	2988	149
2001-2002	288	287	263	285	271	264	258	263	247	230	211	167	3034	149
2002-2003	295	287	280	266	289	268	262	254	269	230	210	172	3082	152
2003-2004	297	299	282	275	268	293	265	266	254	246	206	171	3122	161
2004-2005	304	297	292	284	272	269	289	260	274	235	220	173	3169	162

low point of 194,700 in 1978-79. From there it increased at a rate based mainly on increased births. We can expect that the number of male students will not reach the same level as 1970-71 until 1984-85.

Figure 7.10 also demonstrates the rapid increase in five year olds entering school. In 1970-71 there are only 6,000 five year olds compared to 17,400 six year olds. By 2004-05 this gap has narrowed considerably. There are 22,600 five year olds and 29,300 six year olds.

As an illustration of what happens in any one year, we will use 1993-94 as an example. There are 16,500 five year olds, 24,900 six year olds, etc., until we reach the seventeen year old category which has 10,600 students with 6,6000 who are 18^{+}. The 18^{+} category contains those students in the school system who may be 18 or older.

(2) Figures 7.14 to 7.17 give the projected enrollments for random walk simulation using the parameters derived from the successive subtraction method. These projections are also made by sex for age and grade individually.

Figure 7.14 gives enrollment projections by age for males from 1970-71 to 2004-05. Again we see the rapid climb in five year olds from 1970-71 to 2004-05. However, there is a difference from the comparable results found in Figure 7.10. The projected enrollment now is 216,500 in 1970-71 (greater than in Figure 7.10 by 7,300) which increases slightly in 1971-72 to 217,000. Then

FIGURE 7.14

Projection by Simulation, Age Specific, Male, Parameter Set 3 (Figure 7.8)

(Unit = 100)

Year \ Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970-1971	60	175	183	188	187	191	186	177	174	170	167	158	113	36	2165
1971-1972	60	159	176	183	188	187	191	186	176	174	170	159	126	35	2170
1972-1973	63	147	160	176	183	188	187	191	186	175	174	164	126	34	2154
1973-1974	64	148	147	160	176	183	188	187	191	186	175	169	136	45	2155
1974-1975	66	147	148	147	160	176	183	188	185	191	185	173	141	51	2141
1975-1976	73	148	147	148	147	160	174	183	188	185	191	180	137	58	2119
1976-1977	80	159	148	147	148	147	160	173	182	186	185	187	143	51	2096
1977-1978	84	171	160	148	147	148	147	159	172	182	185	178	156	65	2102
1978-1979	89	175	172	160	148	147	148	147	159	172	182	182	149	71	2101
1979-1980	94	181	176	172	160	148	147	147	147	159	172	177	151	70	2101
1980-1981	98	187	182	176	172	160	148	147	147	147	159	169	148	73	2113
1981-1982	103	191	188	182	176	172	160	148	146	146	147	155	127	75	2116
1982-1983	108	197	192	188	182	176	172	160	148	146	145	146	123	65	2148
1983-1984	113	203	198	192	188	182	175	172	157	148	145	143	105	63	2184
1984-1985	119	209	204	198	192	188	182	174	172	157	147	141	118	65	2266
1985-1986	125	215	210	204	198	191	188	182	173	172	157	141	113	60	2329
1986-1987	131	222	216	210	204	198	191	188	180	172	172	155	112	50	2401
1987-1988	135	227	223	216	210	204	198	190	188	178	172	167	115	43	2466
1988-1989	140	231	228	223	216	210	203	197	190	187	175	170	120	54	2544
1989-1990	145	236	232	228	223	216	209	203	195	189	186	169	130	50	2611
1990-1991	149	239	237	232	228	223	216	209	203	194	188	181	121	57	2677
1991-1992	154	243	240	237	232	228	223	216	208	202	193	186	144	53	2759
1992-1993	162	247	244	240	237	231	228	223	216	208	199	188	143	64	2830
1993-1994	165	251	248	244	240	237	231	227	223	216	208	193	149	69	2901
1994-1995	170	255	252	248	244	240	237	231	227	223	215	200	142	75	2959
1995-1996	175	259	256	252	248	244	240	237	231	226	223	210	144	66	3011
1996-1997	181	263	260	256	252	248	244	238	237	229	225	212	153	54	3052
1997-1998	186	267	264	260	256	252	248	244	238	236	229	221	149	80	3130
1998-1999	191	270	268	264	260	256	252	248	243	237	233	222	167	75	3186
1999-2000	197	274	271	268	264	260	256	252	248	241	235	223	160	72	3221
2000-2001	202	278	275	271	267	263	260	256	252	248	240	231	157	63	3263
2001-2002	208	282	279	275	271	267	263	260	254	252	247	235	163	70	3326
2002-2003	214	286	283	279	275	271	267	262	260	252	252	237	169	75	3382
2003-2004	220	290	287	283	279	275	271	267	262	258	252	247	178	65	3434
2004-2005	226	294	291	287	283	277	274	271	267	258	257	244	167	68	3464

FIGURE 7.15

Projection by Simulation, Age Specific, Female, Parameter Set 4 (Figure 7.9)

(Unit = 100)

Year \ Age	5	6	7	8	9	10	11	12	13	14	15	16	17	18+	TOTAL
1970-1971	56	169	177	179	178	182	173	166	165	162	159	149	98	13	2026
1971-1972	58	148	170	177	179	178	182	173	165	165	161	156	112	8	2032
1972-1973	61	142	148	170	177	179	178	182	173	165	164	156	118	16	2029
1973-1974	60	142	142	148	170	177	179	178	182	172	165	157	117	17	2006
1974-1975	64	137	142	142	148	170	177	179	178	181	171	161	125	17	1992
1975-1976	71	143	137	142	142	148	170	177	179	177	181	166	121	15	1969
1976-1977	78	153	143	137	142	142	148	170	176	179	177	176	127	24	1972
1977-1978	82	165	154	143	137	142	142	148	168	176	179	175	141	23	1975
1978-1979	86	170	166	154	143	137	142	142	148	168	175	175	137	37	1980
1979-1980	90	175	171	166	154	143	137	142	142	148	165	170	145	21	1969
1980-1981	95	180	176	171	166	154	143	137	142	142	148	160	134	22	1970
1981-1982	99	186	181	176	171	166	154	143	137	142	142	147	117	29	1990
1982-1983	105	190	187	181	176	171	166	154	142	137	142	137	113	16	2017
1983-1984	110	197	191	187	181	176	171	166	154	142	137	136	102	21	2071
1984-1985	115	203	198	191	187	181	176	171	166	153	141	135	105	21	2143
1985-1986	120	208	204	198	191	187	181	176	170	166	152	138	105	26	2222
1986-1987	127	214	209	204	198	191	187	181	176	170	166	147	106	17	2293
1987-1988	131	221	215	209	204	198	191	187	181	176	170	160	100	12	2355
1988-1989	136	225	222	215	209	204	198	191	187	181	176	167	117	16	2444
1989-1990	141	228	226	222	215	209	204	198	191	187	181	171	131	15	2519
1990-1991	145	232	229	226	222	215	209	204	198	191	187	174	127	26	2585
1991-1992	150	236	233	229	226	222	215	209	203	198	191	186	109	23	2630
1992-1993	157	239	237	233	229	226	222	215	209	203	198	182	130	27	2707
1993-1994	160	243	240	237	233	229	226	222	215	209	201	194	128	20	2757
1994-1995	165	247	244	240	237	233	229	226	222	215	209	194	132	24	2817
1995-1996	170	251	248	244	240	237	233	229	226	222	215	203	125	28	2871
1996-1997	175	255	252	248	244	240	237	233	229	226	221	213	132	24	2929
1997-1998	180	259	256	252	248	243	240	237	233	226	225	212	134	19	2964
1998-1999	186	263	260	256	252	248	243	240	237	233	226	217	134	20	3015
1999-2000	191	266	264	260	256	252	248	243	240	237	232	217	150	21	3077
2000-2001	196	269	267	264	260	256	252	248	243	240	236	222	147	28	3128
2001-2002	202	273	270	267	264	260	256	252	248	243	238	224	138	22	3157
2002-2003	208	277	274	270	267	264	260	256	252	247	243	230	136	16	3200
2003-2004	213	281	278	274	270	267	264	260	256	251	246	232	155	31	3278
2004-2005	219	285	282	278	274	270	267	263	259	254	250	241	150	27	3319

it starts to drop until it reaches the low point of 209,600 in 1976-77. For the next five years it remains virtually constant and then starts to climb in almost a systematic manner.

The successive subtraction method seems to give consistently higher projections for every year. For example, it is 346,400 in 2004-05 and for the earlier set of parameters it was 328,200 in the same year.

Naturally all the foregoing results must be tempered with caution and should not be taken out of context. Each projection should be used advisedly in conjunction with any other available information.

FIGURE 7.16

Projection by Simulation, Grade Specific, Male, Parameter Set 3 (Figure 7.8)

(Unit = 100)

Grade / Year	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL	GRADUATES
1970-1971	192	210	196	190	199	183	193	178	168	155	145	156	2165	87
1971-1972	186	177	214	195	187	196	198	183	178	155	139	162	2170	96
1972-1973	169	185	179	211	192	187	193	194	194	159	140	151	2154	102
1973-1974	161	170	188	174	210	192	197	190	186	178	150	159	2155	84
1974-1975	164	167	165	181	176	209	197	190	194	169	160	169	2141	95
1975-1976	176	153	167	172	176	175	213	191	190	176	150	180	2119	117
1976-1977	181	174	156	163	165	177	189	213	180	173	163	162	2096	121
1977-1978	193	176	173	156	166	163	180	181	212	167	158	177	2102	104
1978-1979	193	192	178	172	156	170	162	181	185	189	152	171	2101	103
1979-1980	205	188	194	174	165	164	181	144	181	170	165	170	2101	106
1980-1981	214	197	190	195	164	168	175	170	142	167	154	177	2113	121
1981-1982	220	207	194	189	191	167	169	177	160	130	144	168	2116	110
1982-1983	228	215	205	188	199	175	175	168	165	150	116	164	2148	105
1983-1984	235	224	208	204	186	190	185	166	171	150	137	128	2184	105
1984-1985	241	238	208	210	197	188	189	186	166	159	134	150	2266	64
1985-1986	251	231	237	205	207	195	191	190	183	148	146	145	2329	87
1986-1987	256	244	227	242	198	205	202	178	204	157	131	157	2401	89
1987-1988	256	250	243	226	243	192	212	192	184	176	146	146	2466	95
1988-1989	255	260	238	254	220	241	195	206	192	162	158	163	2544	87
1989-1990	265	254	253	241	252	217	236	198	200	178	142	175	2611	99
1990-1991	281	254	244	247	248	250	227	222	195	184	168	157	2677	109
1991-1992	281	280	238	247	249	250	243	225	219	180	167	180	2759	95
1992-1993	275	286	268	240	249	251	251	235	224	201	164	186	2830	120
1993-1994	280	278	277	261	246	252	254	244	223	215	190	181	2901	113
1994-1995	289	279	272	281	258	241	258	241	251	193	190	206	2959	124
1995-1996	297	280	272	270	283	256	254	248	244	217	174	216	3011	111
1996-1997	309	284	277	262	278	277	261	235	256	224	196	193	3052	140
1997-1998	302	309	269	289	254	280	285	247	232	239	205	219	3130	110
1998-1999	312	292	311	265	277	259	284	276	241	217	220	232	3186	158
1999-2000	310	303	298	305	256	281	276	271	269	217	203	232	3221	151
2000-2001	312	306	304	292	304	260	277	264	272	251	199	222	3263	164
2001-2002	320	306	313	299	287	296	273	265	259	249	217	242	3326	132
2002-2003	328	308	310	308	302	281	301	267	255	243	231	248	3382	158
2003-2004	339	316	300	318	300	295	290	297	263	235	215	266	3434	164
2004-2005	346	329	303	299	315	295	292	290	299	240	211	245	3463	178

FIGURE 7.17

Projection by Simulation, Grade Specific, Female, Parameter Set 4 (Figure 7.9)

(Unit = 100)

Year \ Grade	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	TOTAL	[illegible]
1970-1971	182	186	180	197	180	178	161	169	167	151	135	140	2026	105
1971-1972	161	176	184	191	196	186	171	161	168	156	134	148	2032	114
1972-1973	151	160	173	194	192	200	185	174	149	160	137	154	2029	120
1973-1974	146	151	161	179	193	189	208	180	171	139	148	141	2006	126
1974-1975	142	152	152	165	184	189	198	195	181	156	129	149	1992	121
1975-1976	155	148	142	158	164	191	187	210	176	173	135	130	1969	125
1976-1977	166	152	149	150	155	161	199	184	212	159	154	131	1972	100
1977-1978	179	160	150	155	152	155	168	193	181	199	137	146	1975	107
1978-1979	181	177	158	162	150	160	146	169	189	170	173	145	1980	123
1979-1980	186	178	175	159	168	158	154	158	164	167	148	154	1969	119
1980-1981	193	183	180	180	158	168	160	154	156	146	150	142	1970	136
1981-1982	199	190	182	183	185	153	183	155	145	144	122	149	1990	119
1982-1983	203	199	188	184	186	186	159	179	150	135	123	125	2017	128
1983-1984	208	202	199	197	181	191	184	157	175	140	117	120	2071	103
1984-1985	209	215	196	211	201	174	201	175	158	159	123	121	2143	96
1985-1986	218	210	215	200	208	207	187	183	179	145	145	125	2222	91
1986-1987	228	215	216	213	210	209	198	193	179	165	126	141	2293	103
1987-1988	227	234	213	225	215	213	210	191	193	162	150	122	2355	124
1988-1989	234	231	229	224	227	213	210	216	185	178	151	146	2444	100
1989-1990	240	240	220	242	231	216	223	212	204	170	153	168	2519	117
1990-1991	242	245	238	223	241	231	228	213	207	192	152	173	2585	128
1991-1992	252	242	237	253	216	244	241	216	217	187	157	168	2630	155
1992-1993	215	253	243	241	255	219	239	252	207	207	171	169	2707	136
1993-1994	251	253	254	245	246	259	223	232	248	192	176	178	2757	149
1994-1995	259	258	249	258	251	245	258	219	225	236	179	180	2817	152
1995-1996	262	265	251	255	262	250	256	246	218	208	215	183	2871	141
1996-1997	271	261	266	246	264	270	244	260	240	201	184	222	2929	145
1997-1998	272	269	267	262	254	258	276	237	263	226	172	208	2964	177
1998-1999	278	274	271	272	268	249	261	268	243	238	198	195	3015	167
1999-2000	280	280	274	274	280	260	267	253	255	226	226	202	3077	158
2000-2001	282	281	276	289	265	286	266	258	253	239	201	232	3128	154
2001-2002	292	276	288	281	289	261	293	259	253	235	212	218	3157	194
2002-2003	289	294	278	289	292	283	281	282	251	236	210	215	3200	176
2003-2004	299	287	296	286	286	300	274	288	271	231	219	241	3278	162
2004-2005	295	309	284	299	288	290	301	271	281	251	205	245	3319	187

CHAPTER VIII

COMPARING THE MODELS

Throughout this report it has been stressed that projected results should be used with caution, keeping in mind the variabilities and restrictions of the model used. No precise answer can be given to the question, "which model should I use?" and we can only present the potential user with the advantages and limitations of each model to assist him in his choice.

This study concentrated on various classes of the general Markov model after developing several ideas based on the literature as well as exploring many simpler models discussed in Chapter 4. Several projections were made and some eliminated. For example, linear regressions over time seemed invalid due to noticeable decreases in birth and changing migration as well as the poor fit obtained by fitting a straight line through many scattered points. The method of fitting polynomials over time appeared to have some value over the short term but later became unreasonable when certain projected values became negative. The simplest method of eyeballing available data over the short term of one or two years may prove to be every bit as accurate. All of the methods reviewed in the literature have various strengths and weaknesses but none so glaring as their inability to adequately portray student movements so that changes in the system can be detected at an early stage.

Recent advances in methodology place greater attention on the student as an entity in the form of tracing student flows. Most of these advances lean heavily on the Markov model. Why did we use variations or submodels of the Markov model? Mainly to provide some form of verification instead of trying to "test the models," virtually an impossibility in making projections until after the fact. Short (1967) proposes using more than one model as a test. We used several models to provide the user with a means of choosing among detailed models, each having various advantages and disadvantages.

Our time series model has a stochastic element, student enrollment, acted upon by an independent variable, time. Projections were made by age and grade individually. We assumed that the annual number of increases in enrollment for each age or grade is, on the average, the same over an extended period of time. The model is stationary, i.e., the statistical properties of the input and output remain invariant over time, made by forming a successive difference series of the first order.

The only parameters needed for time series projections are those estimated for the model itself by the least squares method. The data required by the model is the enrollment for the previous twenty-five years. No parameters in the form of student transition probabilities, births, deaths, or migration are required.

The time series method is unable to describe student movements between grade-states. The model either fits a function to year and age or year and grade, not year and age and grade. There

are no random fluctuations and the results are smoothed out. This particular model is not easy to change and can only be applied to age or grade specific data.

The time series method itself is quite simply applied and there is very little data collection needed. We recommend investigating other time series forms: second order difference models and moving average models. Over the short term the straight line projections we obtain appear to have merit.

The Markov chain model takes us a step further in the development. We see the flow of the students through the educational system as a Markov process. It is a series of events subject to random fluctuations and proceeding in time so that the next state depends only on the present state. We use parameters that are not products of the model but are internal and external to the system we are studying. The internal parameters are found in the closed part of our system and are in the form of transition proportions representing student flows. Transition proportions or transition probabilities such as pass and failure rates are the most important. External parameters such as birth rates "open" our model exposing it to influences outside of the school system.

We can make the Markov chain model more or less sophisticated depending on our requirements. For example, a subroutine could be developed to input births directly. More complicated revisions, such as making the Markov chain a higher order stochastic model, would make our present model unrecognizable. The function fits year and grade without using age. This could have

been rectified by adopting our parameters to suit this purpose. Aggregate data is fed into the model in providing us with projections by grade.

Important features of the Markov chain model are many. The student flow between levels is described, different parameters can be used, computer time requirements are minimal, the parameters could be varied with time, and it is easily updated.

Our simulation takes the form of a mathematical model portraying a real situation. The random walk simulation is a Markov process that tries to approximate the real situation by following the whole process of student enrollment.

Simulation is a flexible tool. It is not restricted to the Markov property. We can change to a higher order, non-stationary model and apply different sets of parameters to demonstrate the applicability of simulation under varying conditions. It is the only approach which uses individual rather than aggregate data although for the sake of convenience we process one hundred students at a time. We break the model down by age and grade using the age-grade enrollment table as a base. From this we can describe the student flows by age and grade and make projections accordingly. It is the most expensive of the methods in terms of data collection (more parameters are used) and computer time. Cost of computer time could be reduced by using transition proportions rather than generating random numbers. This gets us a step closer to the Markov chain model.

An overriding consideration in the selection of an appropriate method is costs, second only to the needs of the user. Figures 8.1 and 8.2 shows a comparison of the differences in projected enrollment totals for the models. Two results are given for the simulation derived for two different sets of parameters. It may happen that the simpler and more easily applied models yield virtually the same results.

In the preceeding paragraphs the merits and demerits of the individual models have been described and must be fully appreciated before selection of a specific model to use in projection studies. However, probably more pertinent to the users' problem of choosing the most suitable model are the variabilities between the models themselves.

Cost

Amongst the three models time series is the least expensive as it has the lowest data collection costs and only slightly greater costs in computer processing time than the Markov chain which has the lowest cost in this area. The Markov chain has data requirements much more expensive than time series. Simulation has the greatest costs in both categories.

Basic Data

Of the three models only simulation uses individual data and although many years of data are preferred only one year of enrollment data is necessary. Time series and Markov chain use aggregate data with the former needing many years of student

FIGURE 8.1

Projected School Enrollment

Males - 1970-71 to 2004-05

(Unit = 1,000)

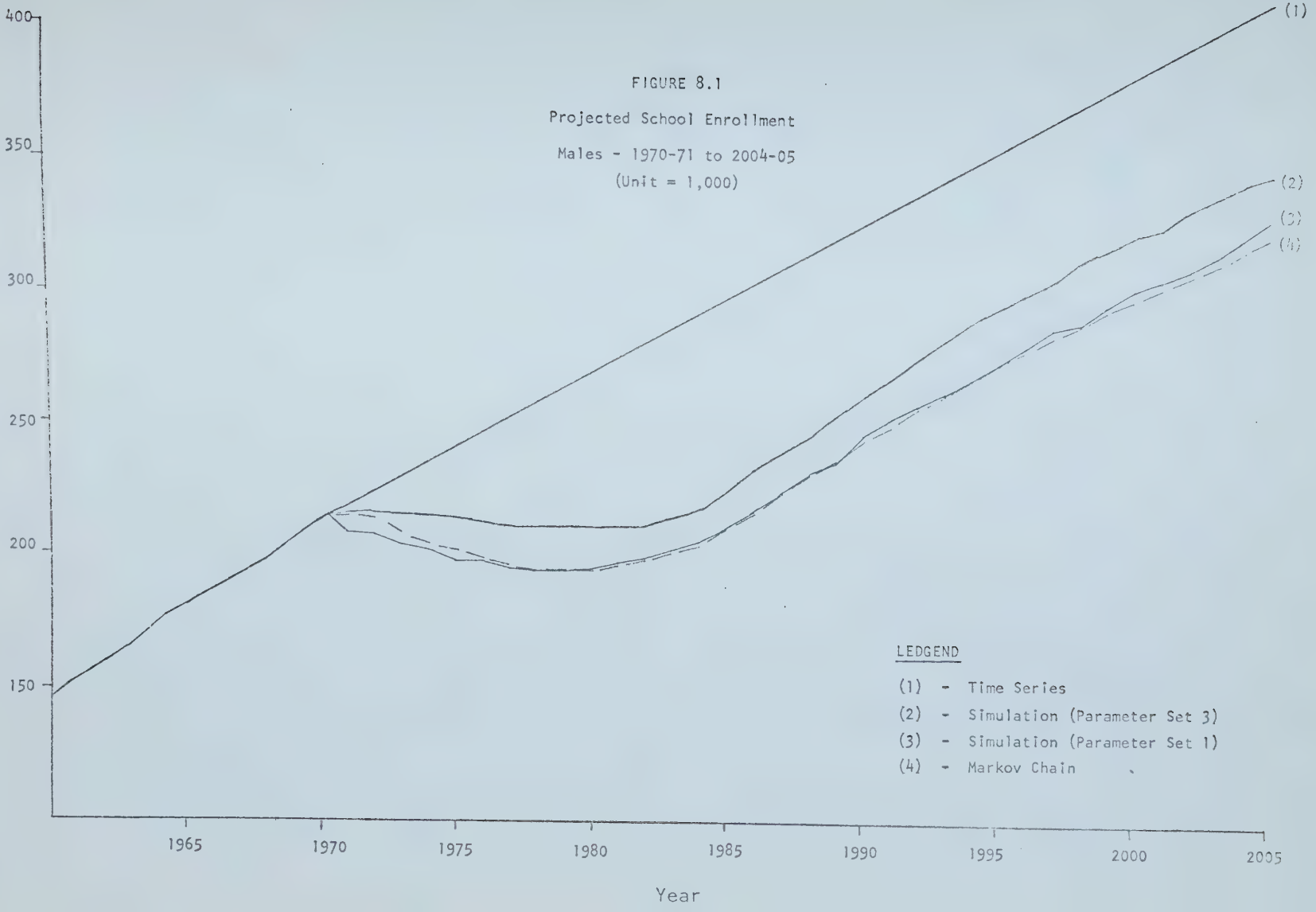
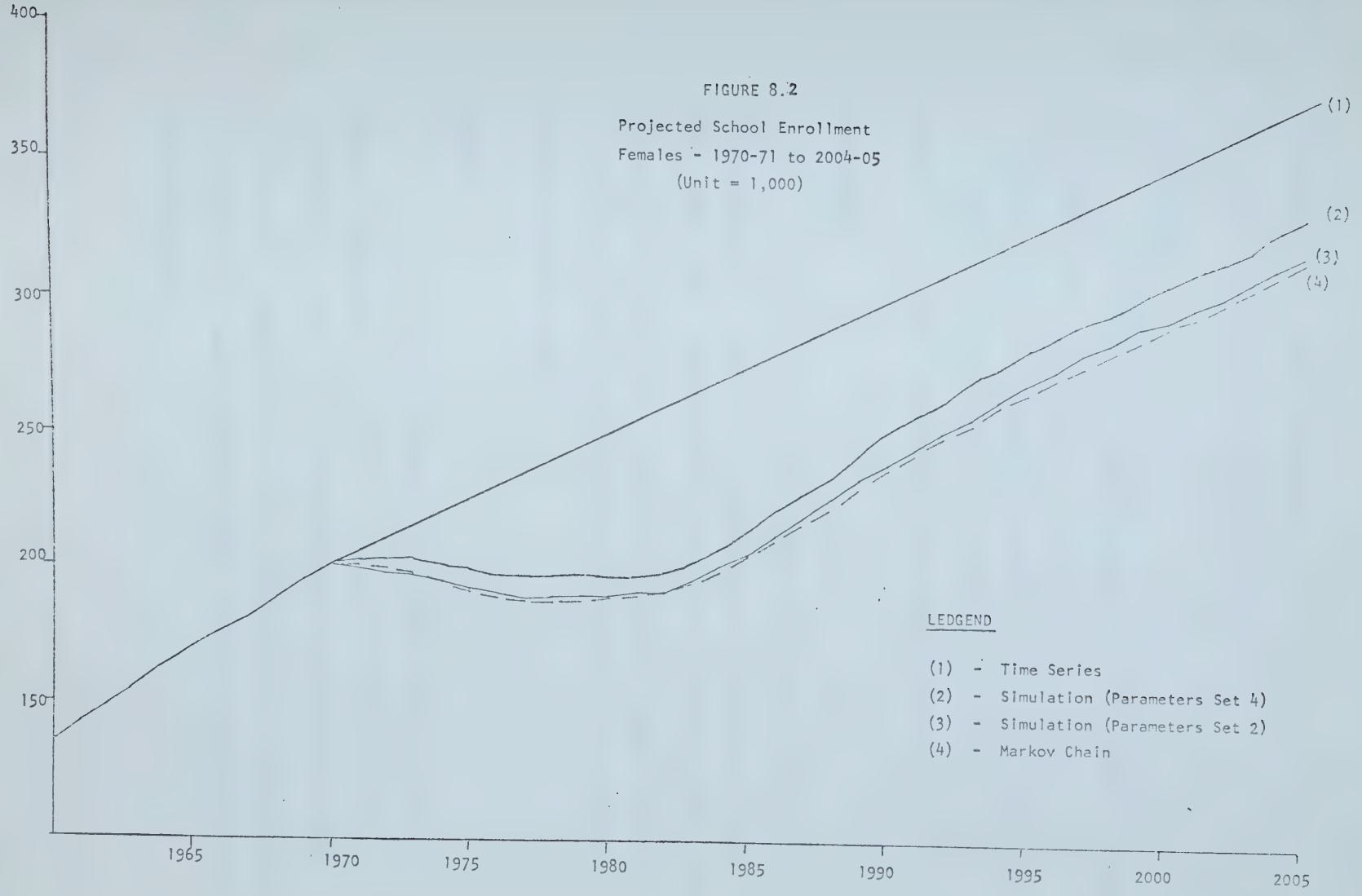


FIGURE 8.2

Projected School Enrollment
Females - 1970-71 to 2004-05
(Unit = 1,000)



enrollment data whereas the Markov chain needs only a few years, although here again many years would be preferred.

Parameters

The time series model generates its own parameters whereas the Markov chain and simulation models must have them supplied. The Markov chain uses internal parameters (transition proportions) and external parameters (births and deaths); simulation needs yet another as it also incorporates the external parameter of migration.

Updating

Time series can only be updated by adding enrollment data and cannot account for changes external to the school system. The Markov chain and simulation models can be easily updated, the Markov chain simply needs to be supplied new information while the simulation can have new information added as well as incorporating other variables.

Output

Time series and Markov chain produce age and grade specific projected enrollments and grade specific projected enrollments respectively. Simulation has more extensive output giving age and grade specific projected enrollments, graduates and dropouts, and it can be easily adapted to output age by grade by year enrollments.

Properties

The time series model fits a function to age and year or grade and year. A Markov chain model fits a function to grade and year without using age. Simulation does not fit a function. Time series has a stationary model due to using a first order difference series. Both the Markov chain and simulation models are also stationary although both, particularly the latter, could be changed to a non-stationary form. All three models are linear.

It is now the responsibility of the potential user to weigh the advantages and disadvantages of each model as it pertains to his problem and to answer the question: "Which model should I use?".

CHAPTER IX

DISCUSSION AND RECOMMENDATIONS

The preceding chapter compared and assessed the projection techniques used in this study as well as discussing other techniques. In addition methodological refinements of the models that might improve projections were discussed. This chapter is also concerned with making more accurate projections through data refinement, parameter refinement and the consideration and/or inclusion of other variables.

9.1 DATA REFINEMENT

While it is possible to formulate a model that theoretically describes a process in great detail with considerable accuracy, the practical application of such a model often requires the collection of detailed information on a number of variables if reliable results are to be obtained. If such information is not readily available or unobtainable, it may be necessary to either use estimated data or to simplify the model by imposing certain assumptions. Both of these alternatives have been used in the present study. For example new entrant totals have been estimated and it has been assumed that all new entrants enter grade one.

Restrictions on the models could be lifted and more reliable projections could be made with additional refinement of the data. Such refinement is less important in the present study,

which has been directed towards the development of a number of school enrollment projection models, than it would be in studies which are more concerned with accurately forecasting student enrollment. Two forms of data refinement which might be useful in accurately forecasting student enrollment involve the use of individual data and the use of cohort data.

9.2 INDIVIDUAL vs AGGREGATE DATA

The school enrollment data used in the present study is aggregated by age, grade, and sex for each school year, that is, the data is grouped. We know, for example, how many grade 1 students in 1959 are five year old males, six year old males, seven year old males and so on. We know also how many grade 2 students in 1960 are six year old males, seven year old males and so on. No precise determination is possible of how many male five year olds in grade 1 in 1959 are among the male six year olds in grade 2 in 1960. Estimations can be made by applying pass-fail rates, dropout rates, and migration rates derived from other sources.

In some instances data is available in individual form, that is, the history of a given individual as he moves through the school system over a period of years is known. Such a history might include many other variables in addition to advancement from grade to grade. Individual data collected over a period of years would allow for accurate determination of pass-fail rates, dropout rates, in- and out-migration rates to mention a few such rates

which could then be applied in making projections. While individual data has these advanatages it also has a number of disadvantages.

The first such disadvantage relates to the quantity of data. Individual data because it involves records for individuals rather than groups of individuals, may result in masses of data which are difficult to handle. Even if one can use the data without aggregating it, the time and cost involved may be uneconomic.

The second problem in using individual data relates to the technique used. While individual data may be useful in generating rates, once the rates are generated or derived from other sources, the input of data into the different projection models may be in individual or aggregate form. In the case of the Markov chain model the appropriate rates are applied to each aggregate or group of students within the system so that fixed proportions of the students progress from state to state at each step. Alternatively, in simulation, each individual in the system moves between states with probabilities corresponding to the calculated rates. Thus the proportion moving between given states at each step is not constant. It should be noted that simulation is more flexible in that it can use either individual or aggregate data.

9.3 AGE COHORTS vs PSEUDO AGE COHORTS

If the raw data is available in individual rather than

aggregate form then it is possible to develop exact age cohorts for input into the selected projection technique. Age cohorts are aggregations of individuals born in a given year who are followed as a group from (in this case) entrance into the school system to final status -- whether pass, failure, dropout, graduation, death or migration. The use of cohort data allows for a more accurate determination of what happens as students progress through the school system.

It should be noted that data is not frequently available in individual or cohort form. In the present case it was not possible to develop age cohorts, although pseudo age cohorts were used. To refer to the earlier example a pseudo age cohort would be those students who were five years old in grade 1 in 1959, six year olds in grade 2 in 1960, seven year olds in grade 3 in 1961, and so on. This is an inexact cohort in that some six year olds in grade 2 may have entered at grade 2 or have migrated into the system at that level rather than appearing as five year olds in grade 1.

When data is input into a model using age, grade, and year it is possible to develop the pseudo age cohorts described above. When the data input is broken down only by grade and year, it is not possible to make use of the cohort year concept. In this case it might be said that the data is aggregated by regular year rather than cohort year. Both of the above forms of aggregation were used, the first in the simulation model the latter in the Markov chain model.

Because cohort data is built from individual data it is often difficult to obtain. Cohort data is no more difficult or expensive to use than any other form of aggregated data. The determination of rates is a relatively easy task if cohort data is available. In conclusion the cohort data has some of the advantages of individual data as well as the advantage of aggregate data.

In the text, pseudo age cohorts are referred to by the term cohorts.

9.4 PARAMETER REFINEMENTS

The accuracy of the predicted estimates of school enrollment for the future depends not only on the validity of the models and the data, but also upon the accuracy of the parameters. Each of the projection models used in this study requires parameters, some of which must be estimated by other methods or from other sources.

The most important parameters are those associated with the number of students entering school. Estimates of the number of students entering school may be made from birth projections, which in turn are frequently associated with general population projections. This study has used the birth projections of the Alberta Bureau of Statistics, which were derived by simplified methods that may not be very reliable, but the estimates were readily available. The intent of this report was not to concentrate on the accuracy of the parameters used, but rather to make

use of the best information already available. There remains a great need for intensive study of birth projections, since they are of great value to research and planning in diverse areas.

Of lesser importance to long-range forecasting are estimates of mortality and migration. Mortality is a very small and relatively constant factor, while migration is a larger and more variable factor. Although an attempt was made to use migration as one of the parameters in the present study, the attempt was abandoned because of the variability in past migration rates and the difficulty in obtaining accurate estimates. The influence of migration is small enough that any improvement in the projections is more than offset by the error introduced by the inadequacy of the estimates. Migration might be a useful parameter to include if individual data or more accurate estimates of aggregate data were available. Once the extensive migration data from the 1971 census have been analyzed, more accurate migration information will be available.

Estimates of passing, failing, and dropping out of school were made from the enrollment data base by the method described in Appendix B. Extensions and refinements of this method or other methods could provide more accurate estimates of these parameters which are internal to the school system. Also, as individual data on school enrollment may become available in the future, greater precision in estimates of these parameters would be possible.

9.5 OTHER VARIABLES

In the previous section, migration was discussed as a parameter that might be included in future projections. This is but one of many variables that might be considered in attempts to make more refined projections. Such factors as socio-economic status and intelligence might be crucial in determining pass-fail rates, dropout rates, and high school graduation rates and rates of continuation into higher education. A consideration of the proportions of students in different educational streams (i.e., vocational and academic) or different disciplines would be of some value particularly for projections in regard to post-secondary education. Projections with regard to regional distributions or sectarian -- non-sectarian distributions might also be a useful refinement. Which of these variables one might include will depend on data availablility, time and cost limitations, technique limitations, and the purpose of the projections.

Factors such as the above can be foreseen to have an effect on the accuracy of projections. Other factors that affect the accuracy of projections often cannot be considered for inclusion because their influence cannot be foretold. Changes such as lowered age of entry into the educational system (i.e., kindergarten) or completely free post-secondary education can have considerable ramifications but cannot generally be taken into account in advance. Until the event has occurred it cannot be foreseen and cannot be included in the projections. Other factors such as sudden rises in unemployment or increased draft calls may affect enrollments

remarkably and unexpectedly at certain levels in the system.

9.6 CHOOSING AMONG TECHNIQUES

Three major techniques have been presented for producing school enrollment projections. All have their roots as Markov models but as noted in the previous chapter have distinctive characteristics. The review of the literature noted methods of a less complex nature -- grade-cohort and enrollment ratio -- that could easily be applied to our situation. Some of these were tried as background study for our work.

It is not possible to state definitively what technique individuals or groups contemplating doing projections might use. This choice depends upon a number of factors: How much money is available? How far ahead does one wish to project? How much data is required? How much data is available? These and many other questions must be settled before selecting a method.

Some minimum necessities are a must. Obvious ones are the collection of enrollment data for at least the previous five years and a review of any major administrative or policy changes that have taken place in the system during that period. For projections over the short term "quick and dirty" methods are advocated if only rough approximations are required. The "dirty" methods are crude mathematical procedures consisting mainly of eyeball projection or straight line extensions of existing enrollment data. Even these require extensive data collection,

although they are cheap in manpower and time and, therefore, cost, over the longer term more formal and expensive but more accurate methods requiring lengthy and detailed data collection as well as the hiring of competent people are a must.

There is one area in which a concrete recommendation can be made. School systems should seriously consider developing a computer based projection model that can be readily updated to meet changes in the system. The initial costs of such a program may seem great (actually this cost would be a relatively small part of total administrative expenditures) but the assistance a well constructed model would give in educational planning more than compensate for the initial outlay. Initial costs of computer based models is further compensated for by the ease and small cost in adding information and updating the model.

APPENDIX A

CONVERSION OF THE ALBERTA SCHOOL ENROLLMENT DATA 1945-46 to 1969-70

A.1 INTRODUCTION

Because the age distribution of students within each grade changes throughout the course of a school year the detection of trends in enrollments by age and by grade requires student enrollment figures to conform to a uniform reporting date for all school years under investigation. Otherwise, any results obtained from the data will be confounded by the time factor. An accurate conversion of an enrollment table to a new reporting date requires detailed birth date information for all students. Since, as is the present case, such information is generally not readily available or is impossible to obtain, a method for estimating the change in an enrollment distribution from one reporting date to another is required. The method used for the present study to convert the Alberta school enrollment data to conform to a uniform reporting date is given below. Although this method is somewhat arbitrary a standard procedure for converting the data is not available at the present time.

A.2 THE CONVERSION PROCEDURE

For the school years 1945-46 to 1969-70 the Alberta student

enrollment figures were reported according to their distribution at the following dates:

(i) 1945-46 to 1960-61 - Figures given as of June 1.

(ii) 1961-62 to 1963-64 - Figures given as of October 31.

(iii) 1964-65 to 1969-70 - Figures given as of September 30.

The policy for Alberta schools over the last six years, 1964-65 to 1969-70, has been to report student enrollment totals as of September 30. This date has been taken as our standard as we expect this to be the recording date over the near future. The three steps of the conversion procedure are described below.

(a) *Age Total Adjustment*

For the period 1945-46 to 1960-61 student enrollments were tabulated as of June 1, a delay of eight months from the September 30 reporting date. Investigation of the distribution of births for each year revealed that the number of births per month was practically constant, therefore, the adjusted age totals for the age groups 7 to 21 were obtained by setting each adjusted age total equal to one-third the original age total plus two-thirds the original age total for the next age group. Eight months is two-thirds of a year, hence, eight months earlier one-third of the students in a particular age group would be expected to remain in that age group and two-thirds would be expected to be in the previous age group. If $x_{i.}$ is the number of students aged i years on June 1, where the dot indicates summation over grade, the estimate for the number of students $y_{i.}$ aged i years on September 30 is:

$$\hat{y}_{i.} = \frac{1}{3} x_{i.} + \frac{2}{3} x_{i+1.} \quad .$$

Three exceptions were made in making the adjustments. These were:

(i) The 21 year old age total was simply reduced by two-thirds as no 22 year olds appeared in the tables.

(ii) The number of 6 year olds was equated to the number of 7 year olds.

(iii) The number of 5 year olds was obtained by subtracting the adjusted age totals for age groups 6 to 21 from the total for all ages.

The adjustments for the 5 and 6 year old totals were made in the above manner as a number of prospective students in these age groups for the population will be entering the school system at a later date, therefore, the student age distributions in these cases will not be representative of the population age distributions and will tend to be shifted towards the later months of the year. For example, 5 year old students will usually be nearly 6 years old. The total for 6 year olds was equated to the 7 year old total as these totals were found to be nearly equal in the enrollment tables for the school years 1964-65 to 1969-70. The remaining students were placed in the 5 year old category as the total enrollment is not affected by the adjustments.

For the school years 1961-62 to 1963-64 the procedure was the same as that outlined above except the one-third and two-thirds adjustments were replaced by adjustments of one-twelfth and

eleven-twelfths respectively. The required correction being just one month for this period.

(b) *Iteration*

Adjusted age by grade totals were obtained from the adjusted age totals and the original grade totals by iteration. This iteration procedure adjusts the original age-grade totals so that their sum by grade equals the adjusted age totals and their sum by age equals the original grade totals. The procedure consists of repeated multiplication of the age-grade totals by the ratio of the required to the observed age and grade totals until their sum by age and by grade equals the desired grade and age totals, respectively.

The first step of the iteration procedure is to multiply each age-grade total by the corresponding ratio of its adjusted age total to its original age total. Now, the age-grade totals will sum by grade to the required adjusted age totals, but their grade totals will generally differ from the original totals. Thus, the second step is to multiply each new age-grade total by the corresponding ratio of the original grade total to the observed grade total. The observed age totals will now differ from the adjusted age totals. Therefore, steps one and two are performed once again using the observed age totals and grade totals. If this is done repeatedly the age-grade totals summed by grade and by age will converge to the required age and grade totals, respectively.

In equation form each iteration produces adjusted age-grade totals given by

$$y_{ij} = Rx_{ij} , \qquad \begin{matrix} i = 5,6,\ldots,21 \\ j = 1,2,\ldots,12 \end{matrix}$$

where y_{ij} and x_{ij} are the iterated and observed age-grade totals, and R is the ratio of the required to the observed age or grade total, as the case may be. For example, the age 5 totals for each grade, adjusted by iteration on age totals, are given by:

$$y_{5j} = \frac{y_{5.}}{x_{5.}} x_{5j} , \qquad j = 1,2,\ldots,12$$

where $y_{5.}$ and $x_{5.}$ are the required and the observed age 5 totals, respectively.

For the Alberta data this iteration procedure was performed on each enrollment table with age totals adjusted to the September 30 reporting date producing revised tables with age-grade totals that summed by grade and by age to the required age and grade totals, respectively. Prior to the actual application of the procedure all age-grade totals less than six were set equal to zero to prevent diproportionate age-grade totals resulting from cases where the original age totals were very small. This applied particularly to tables for the earlier years where the original number of 5 year old students was very small due to the late reporting date and was greatly increased by the age total adjustments of (a).

(c) *Regression*

After the adjustments outlined in (a) and (b) had been completed small jumps in the enrollment figures could still be detected in the enrollment tables between school years 1960-61 to 1961-62 and 1963-64 to 1964-65. To smooth out these jumps a regression equation was fitted to the totals for each age-grade combination.

The regression equation has the form

$$T_y = A_1X_{1y} + A_2X_{2y} + A_3X_{3y} + A_4X_{4y} + A_5X_{5y} + E_y$$

$$y = 1,2,\ldots,25 \quad ,$$

where,

T_y = age-grade total for year y

X_{1y} = 1 for all y

X_{2y} = Year

X_{3y} = Year squared

X_{4y} = 0 except $X_{4,17} = X_{4,18} = X_{4,19} = 1$

X_{5y} = 0 except $X_{5,20} = X_{5,21} = \ldots = X_{5,25} = 1$

E_y = Residual

The regression coefficients A_1, A_2, and A_3 account for the underlying trend in the age-grade totals, A_4 accounts for the jump occurring at school year 1961-62, and A_5 accounts

for the jump occurring at school year 1964-65.

The estimates $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_5$ of the equation parameters $A_1, A_2, \ldots, A_5$, obtained by the least squares procedure, were used to obtain adjusted age-grade totals given by:

$$\hat{T}^{*}_{y} = \hat{A}_1 X_{1y} + \hat{A}_2 X_{2y} + \hat{A}_3 X_{3y} + \hat{A}_5 + \hat{E}_y \quad , \qquad (A.2)$$

$$y = 1,2,\ldots,25 \ .$$

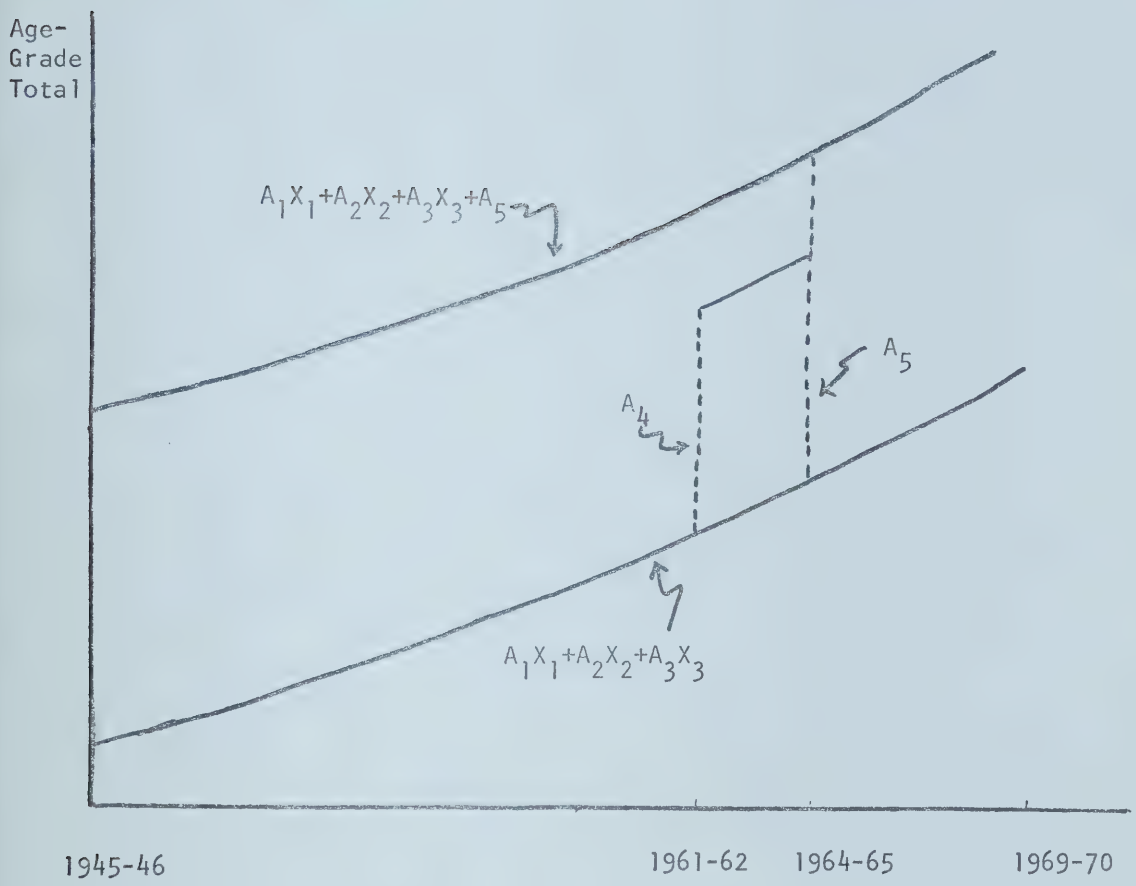
This produced age-grade totals that conformed to the underlying trend given by:

$$\hat{A}_1 X_{1y} + \hat{A}_2 X_{2y} + \hat{A}_3 X_{3y}$$

plus the increment $\hat{A}_5$. Thus the age-grade totals for the school year 1964-65 to 1969-70 were reproduced exactly and the age-grade totals for the proceeding years were made to conform to this period by the increment A_5, as illustrated in Figure A.1.

FIGURE A.1

Regression Model for the Adjusted
Age-Grade Totals



The regression procedure produced a few small negative age-grade totals in some of the tables, and in general the observed grade totals differed slightly from the original grade totals. To correct for this all negative age-grade totals were set equal to

zero and then the age-grade totals for all tables were multiplied by the ratio of its original to its observed grade total. The resulting tables were taken as the required tables converted to the September 30 reporting date.

APPENDIX B

SECTION I: PARAMETER ESTIMATION

B.1 PARAMETER ESTIMATION

Some of the parameters required by the models in this report are estimated from data external to the school system. This is true of birth rates, migration rates, and the like. However, parameters such as pass and failure rates may be estimated from aggregate school enrollment data by the method discussed here.

The method involves successive subtractions of enrollment data at two consecutive ages in an age-grade cohort table. It attempts to account exactly for the observed cell frequencies in terms of elementary processes of passing, failing, entering school, and leaving school. The processes have the following order of precedence: passing has priority over entering school, and failing has priority over leaving school. Passing and failing have the same level of priority. If the data can be explained by a higher priority process, then a lower priority process is assumed to be absent. Skipping grades (accelerating) is not one of the assumed processes. An independent estimate of the number of students leaving school from the highest grade, including both graduates and dropouts, is necessary to start the procedure.

B.1.1 Example of a Portion of a Cohort Table

Cohort Table for Males 18 in 1964-65

Age	Grade	10	11	12
...	...	...	...	...
15	...	5055	4325	1090
16	...	2620	4042	3671
...	...	...	...	...

Note that the row for age 15 is from enrollment data for the year 1961-62, while the age 16 row is for year 1962-63, since the entire table consists only of the age cohort of males who are 18 years old in 1964-65. A cell is a position in the table, which has a value equal to the table entry at that position. For example, the value of the cell at age 16 and grade 11 is 4042. A more general portion of a cohort table, illustrating the notation, follows:

Age	Grade	g-2	g-1	g
...	...	...	...	...
a-1	...	(a-1,g-2)	(a-1,g-1)	(a-1,g)
a	...	(a,g-2)	(a,g-1)	(a,g)
...	...	...	...	...

The purpose of the method is to estimate transition frequencies in a cohort table from age a-1 to age a the following year, for all grades.

B.1.2 Notation and Definitions

(1) a,g denotes the cell at age a and grade g of a cohort table: e.g., 16,11 is the cell at age 16 and grade 11,

(2) students not present in the schools are denoted as being in cell 0.

(3) (a,g) denotes the number of students in a cell at age a and grade g of a cohort table. For the above example, (16,11) has the value 4042.

(4) (a-1,g → a,g') denotes the estimated number of students who went from cell a-1,g to cell a,g' the following year. This is an estimated transition frequency. For example, (15,11 → 16,12) denotes the estimated number of students who went from age 15 grade 11 to age 16 grade 12 the following year (that is, they passed from grade 11 to grade 12).

(5) All subtractions, denoted by -, are of a special type. Negative results are not permitted. Instead, a potentially negative result is set equal to zero.

e.g., 3-2 = 1

but 2-3 = 0 .

The result of the latter example is said to be potentially negative, because it would have been negative except for the rule. It follows immediately from this rule that the following algebraic function

is bivalued:

$$X - (X-Y) = X \quad \text{if} \quad X \leq Y$$
$$= Y \quad \text{if} \quad X > Y$$

Also, the following identities hold:

(a) $Y - [X - (X-Y)] = Y-X$

(b) $(X-Y) + [X - (X-Y)] = X$

(c) $(Y-X) + [X - (X-Y)] = Y$.

(6) The more common transition frequencies are named and illustrated as follows:

(a) Passing $(a-1,g-1 \rightarrow a,g)$

(b) Failing $(a-1,g-1 \rightarrow a,g-1)$

(c) Leaving School $(a-1,g-1 \rightarrow 0)$

(d) Entering School $(0 \rightarrow a,g)$

(7) The lowest grade is denoted by 1.

B.1.3 Method

The method estimates transition frequencies, from age $a-1$ to age a, that will exactly explain the observed data. The procedure accounts for individual cell values, starting initially from the highest grade g.

(1) The process begins with an independent or external estimate of the transition frequency $(a-1,g \rightarrow 0)$ for leaving school from the highest grade g. In the example, grade $g = 12$ is the highest grade, and suppose that $(15,12 \rightarrow 0)$ has been

estimated to be 900.

(2) The remaining students must have failed, $(a\text{-}1,g \rightarrow a,g) = (a\text{-}1,g) - (a\text{-}1,g \rightarrow 0)$. In the example,

$$(15,12 \rightarrow 16,12) = (15,12) - (15,12 \rightarrow 0)$$

$$(15,12 \rightarrow 16,12) = 1090 - 900 = 190$$

Cell value $(a\text{-}1,g)$ has now been accounted for.

(3) Students (a,g) are assumed to have originated in three ways:

(a) failing from cell $a\text{-}1,g$ as already estimated

(b) passing from cell $a\text{-}1,g\text{-}1$

(c) entering from outside the school system.

Alternative (a) has already been estimated as the value $(a\text{-}1,g \rightarrow a,g)$, so only the quantity A, where

$$A = (a,g) - (a\text{-}1,g \rightarrow a,g)$$

remains in cell a,g to be accounted for by processes (b) and (c). Quantity A is the number of students in cell a,g that did not get there by failure from cell $a\text{-}1,g$.

Cell value $(a\text{-}1,g\text{-}1)$ is the maximum number of students that could have passed to cell a,g. If this quantity is less than the amount A remaining to be accounted for in cell a,g then the difference is assumed to be due to process (c). Thus the number of new entrants is estimated to be

$$(0 \rightarrow a,g) = A - (a\text{-}1,g\text{-}1) \quad .$$

The reader should keep in mind that negative results are not permitted in subtraction.

The number of students passing from a-1,g-1 is then estimated as the amount still remaining to be accounted for in cell a,g:

$$(a\text{-}1,g\text{-}1 \rightarrow a,g) = A - (0 \rightarrow a,g)$$
$$= A - [A - (a\text{-}1,g\text{-}1)] \quad .$$

It is usually the case that (a-1,g-1) exceeds A, so that new entrants $(0 \rightarrow a,g)$ are absent and the frequency of passing is A. However, if A exceeds (a-1,g-1), then the difference is equal to the number of new entrants, and all students in cell a-1,g-1 passed to a,g. Passing thus takes precedence over entering.

For the example, the quanities are as follows:

$$A = (16,12) - (15,12 \rightarrow 16,12)$$
$$= 3671 - 190$$
$$= 3481$$

$$(0 \rightarrow 16,12) = A - (15,11)$$
$$= 3481 - 4325$$
$$= 0$$

$$(15,11 \rightarrow 16,12) = A - (0 \rightarrow 16,12)$$
$$= 3481 - 0$$
$$= 3481$$

Thus, of the 3671 16 year olds in grade 12 it is estimated that 190 arrived there by failing grade 12 the previous year, 3481 passed from grade 11, and there are no new entrants.

(4) Cell a-1,g-1 is analyzed next. It is assumed that three processes account for where these students went the following year:

(a) passing to cell a,g as already estimated

(b) failing to a,g-1

(c) leaving the school system.

Alternative (a) has already been estimated as the value $(a\text{-}1,g\text{-}1 \rightarrow a,g)$, so only the quantity B, where

$$B = (a\text{-}1,g\text{-}1) - (a\text{-}1,g\text{-}1 \rightarrow a,g)$$
$$= (a\text{-}1,g\text{-}1) - A$$

remains in cell a-1,g-1 to be accounted for by processes (b) and (c). Quantity B is the number of students in cell a-1,g-1 that did not pass to cell a,g.

Cell value (a,g-1) is the maximum number of students that could have failed into cell a,g-1. If this quantity is less than the amount B remaining to be accounted for in cell a-1,g-1 then the difference is assumed to be due to process (c). Thus the number of students leaving school is estimated to be

$$(a\text{-}1,g\text{-}1 \rightarrow 0) = B - (a,g\text{-}1) \quad .$$

The number of students failing from a-1,g-1 to a-1,g is then estimated as the amount still remaining to be accounted for

in cell a-1,g-1:

$$(a\text{-}1,g\text{-}1 \rightarrow a,g\text{-}1) = B - (a\text{-}1,g\text{-}1 \rightarrow 0)$$
$$= B - [B - (a,g\text{-}1)] \quad .$$

It is usually the case that (a,g-1) exceeds B, so that dropouts are absent and the frequency of failing is B. However, if B exceeds (a,g-1), then the difference is equal to the number leaving school, and all students in cell a,g-1 failed from a-1,g-1. Failing thus takes precedence over leaving school.

For the example, the quantities are as follows:

$$B = (15,11) - (15,11 \rightarrow 16,12)$$
$$= 4325 - 3481$$
$$= 844$$

$$(15,11 \rightarrow 0) = B - (16,11)$$
$$= 844 - 4042$$
$$= 0$$

$$(15,11 \rightarrow 16,11) = B - (15,11 \rightarrow 0)$$
$$= 844 - 0$$
$$= 844.$$

Thus, of the 4325 15 year olds in grade 11, it is estimated that 3481 passed, none dropped out, and 844 failed.

(5) Students (a,g-1) can now be analyzed as was (a,g) in step (3), followed by repetition of (4) to account for (a-1,g-2), etc.

(6) Thus steps (3) and (4) are repeated with decreasing values of g, until $g = 1$ in step (3). Then alternative (3) (a) is not available; instead students are assumed to be entering school for the first time:

$$(0 \rightarrow a,1) = (a,1) - (a-1,1 \rightarrow a,1) \quad ,$$

and the estimation procedure is completed. The estimated transition frequencies can be readily converted to transition rates or proportions by dividing by the appropriate cell frequencies.

B.1.4 Alternative Methods

As an alternative to the above method, individual cell values could be analyzed in turn starting at the lowest grade first. The process would then begin with an independent or external estimate of the transition frequency $(0 \rightarrow a,1)$.

Either method can be used to analyze data that is summed over age, i.e., not age-dependent. In this case, data is not arranged in cohort form but simply year-by-year. Symbol (a,g) can then refer to the data by grade for a given year, while $(a-1,g)$ are the corresponding grade totals for the previous year. The resulting estimates of transition frequencies are then year-specific but not age-specific.

Other methods of parameter estimation, and the resulting estimates, are mentioned in this report where appropriate. The authors have relied upon published estimates for births and migration, and have extended these as necessary.

APPENDIX B

SECTION II: ESTIMATION OF NEW ENTRANTS

B.2 ESTIMATION OF NEW ENTRANTS

The accuracy of projecting school enrollments relies heavily on the estimation of new entrants to the school system at the beginning of each school year. The following is a description of the method and assumptions used in the present study to estimate new entrants, those students enrolling in grade one for the first time, to the Alberta School System for the school years 1970-71 to 2004-05.

The method is based on the following assumptions:

(i) All new entrants enter grade 1.

(ii) Students enter school for the first time at age 5, 6, 7, or 8. No new entrants occur for subsequent age groups.

(iii) Ninety-five per cent of those born in a given year will eventually enter the school system.

(iv) The new entrant rates are the same for males and females.

(v) The new entrant rates are constant over time from year to year for ages 7 and 8.

The assumptions are based on analysis of actual new entrant data for school years prior to 1970-71, and were made to simplify the problems of new entrant estimation to manageable proportions. Analysis of the data revealed approximately 95 per

cent of those born in a given year eventually enter school and that little difference exists between male and female entrance rates. The five percent loss to the system is mainly due to mortality and health reasons. The new entrants for ages 7 and 8 were found to account for less than two per cent of the new entrant total for all ages and the total entering school for the first time beyond age 8 was found to be negligible. Thus, the age 7 and 8 entrance rates were assumed constant, and the rates for the older age groups were taken to be zero.

The first step in estimating new entrants was to obtain birth projections for the years 1970 to 2005 and to combine these projections with the actual birth totals for the years 1962 to 1969. The actual birth totals were obtained from the Dominion Bureau of Statistics (1970) and the projections were derived on the basis of population forecasts published by the Alberta Bureau of Statistics (1968), which gave birth projections, at five year intervals, for the years 1971 to 1986. Interpolation and projection were carried out on these forecasts by extending a line through the plotted points to obtain birth estimates for the individual years 1970 to 2005. Although these birth estimates may not be very reliable, they are the only available birth projections for Alberta at this time. (The Oil and Gas Conservation Board has projected estimates of total population, but birth projections are not reported.)

Let,

B_t = the number of births in year t.

$p_{i,t+i}$ = the proportion of those born in year t who enter grade 1 for the first time in year t+i, aged i years.

$E_{i,t+i}$ = the number of new entrants to grade 1 aged i years in year t+i.

Then,

$$E_{5,t+5} = p_{5,t+5} \cdot B_t$$

$$E_{i,t+i} = p_{i,t+i} \cdot (1 - p_{5,t+5}) \times \ldots \times (1 - p_{i-1,t+i-1}) \cdot B_t \qquad (B.2.1)$$

$$i = 6,7,8,$$

and the proportion of the births B_t who enter grade 1 is

$$K_t = [p_{5,t+5} \cdot B_t + p_{6,t+6} \cdot (1-p_{5,t+5}) \cdot (1-p_{5,t+5}) \cdot (1-p_{6,t+6}) \cdot B_t + p_{8,t+8}(1-p_{5,t+5}) \cdot (1-p_{6,t+6}) \cdot (1-p_{7,t+7}) \cdot B_t] \div B_t$$

$$= p_{5,t+5} + (1-p_{5,t+5})[(p_{6,t+6} + p_{7,t+7}(1-p_{6,t+6}) + p_{8,t+8} \cdot (1-p_{6,t+6}) \cdot (1-p_{7,t+7})] \quad . \qquad (B.2.2)$$

The new entrant rates for the school year 1967-68 were estimated from the new entrant data given in the Alberta Department of Education Annual Reports. These estimates were used as the basis

FIGURE B.2.1

New Entrant Rates for School Years
1967-68 to 2004-05

Year	Age 5	6	7	8
1967-68	0.26	0.91	0.22	0.02
1968-69	0.28	0.91	0.22	0.02
1969-70	0.31	0.91	0.22	0.02
1970-71	0.37	0.90	0.22	0.02
1971-72	0.40	0.89	0.22	0.02
1972-73	0.42	0.89	0.22	0.02
1973-74	0.43	0.89	0.22	0.02
1974-75	0.44	0.89	0.22	0.02
1975-76	0.45	0.88	0.22	0.02
1976-77	0.46	0.88	0.22	0.02
1977-78	0.47	0.88	0.22	0.02
1978-79	0.48	0.88	0.22	0.02
1979-80	0.49	0.87	0.22	0.02
1980-81	0.50	0.87	0.22	0.02
1981-82	0.51	0.87	0.22	0.02
1982-83	0.52	0.87	0.22	0.02
1983-84	0.53	0.86	0.22	0.02
1984-85	0.84	0.86	0.22	0.02
1985-86	0.55	0.86	0.22	0.02
1986-87	0.56	0.85	0.22	0.02
1987-88	0.57	0.85	0.22	0.02
1988-89	0.58	0.85	0.22	0.02
1989-90	0.59	0.84	0.22	0.02
1990-91	0.60	0.84	0.22	0.02
1991-92	0.61	0.84	0.22	0.02
1992-93	0.62	0.83	0.22	0.02
1993-94	0.63	0.83	0.22	0.02
1994-95	0.64	0.82	0.22	0.02
1995-96	0.65	0.82	0.22	0.02
1996-97	0.66	0.81	0.22	0.02
1997-98	0.67	0.81	0.22	0.02
1998-99	0.68	0.80	0.22	0.02
1999-00	0.69	0.79	0.22	0.02
2000-01	0.70	0.79	0.22	0.02
2001-02	0.71	0.78	0.22	0.02
2002-03	0.72	0.77	0.22	0.02
2003-04	0.73	0.76	0.22	0.02
2004-05	0.74	0.75	0.22	0.02

for the estimation of new entrant rates for the school years 1967-68 to 2004-05 given in Figure B.2.1.

The new entrant rates given in Figure B.2.1 were calculated as follows. The age 7 and 8 rates were assumed constant from year to year and equal to the 1967-68 estimates and the age 5 rates were calculated from the actual enrollment data for school years 1967-68 to 1970-71. For the age 5 entrance ratio an annual increase of three, two, and one per cent was assumed for the school years 1971-72, 1972-73, and 1973-74, respectively, and an annual increase of one per cent was assumed for each subsequent year. These latter assumptions are somewhat arbitrary and more accurate estimates might be obtained on the basis of changes forecasted for grade 1 admittance requirements such as the recent change in the age requirement of six years by December 31 to February 28. Having determined the age 5, 7, and 8 year old entrance rates, and assuming that 95 per cent of those born each year eventually enter school, the age 6 entrance rates were determined from equation (B.2.2) in the following manner:

By equation (B.2.2),

$$.95 = p_{5,t+5} + (1-p_{5,t+5})[(p_{6,t+6} + p_{7,t+7})(1-p_{6,t+6}) + p_{8,t+8}(1-p_{6,t+6})(1-p_{7,t+7})] .$$

Considering the term $p_{8,t+8} \cdot (1-p_{5,t+5})(1-p_{6,t+6})(1-p_{7,t+7})$ to be negligible, we have,

$$.95 = p_{5,t+5} + (1-p_{5,t+5})[p_{6,t+6} + p_{y,t+7}\ (1-p_{6,t+6})] \ ,$$

or

$$p_{6,t+6} = \frac{.95 - [p_{5,t+5} + p_{7,t+7}(1-p_{5,t+5})]}{(1-p_{5,t+5})(1-p_{7,t+7})} \ . \qquad \text{(B.2.3)}$$

Equation (B.2.3) provides an estimate of $p_{6,t+6}$ for each year.

The new entrant estimates that were obtained are tabulated in Figure B.2.2, and the birth projections along with the new entrant estimates for the combined male and female case are shown graphically in Figure B.2.3.

FIGURE B.2.2

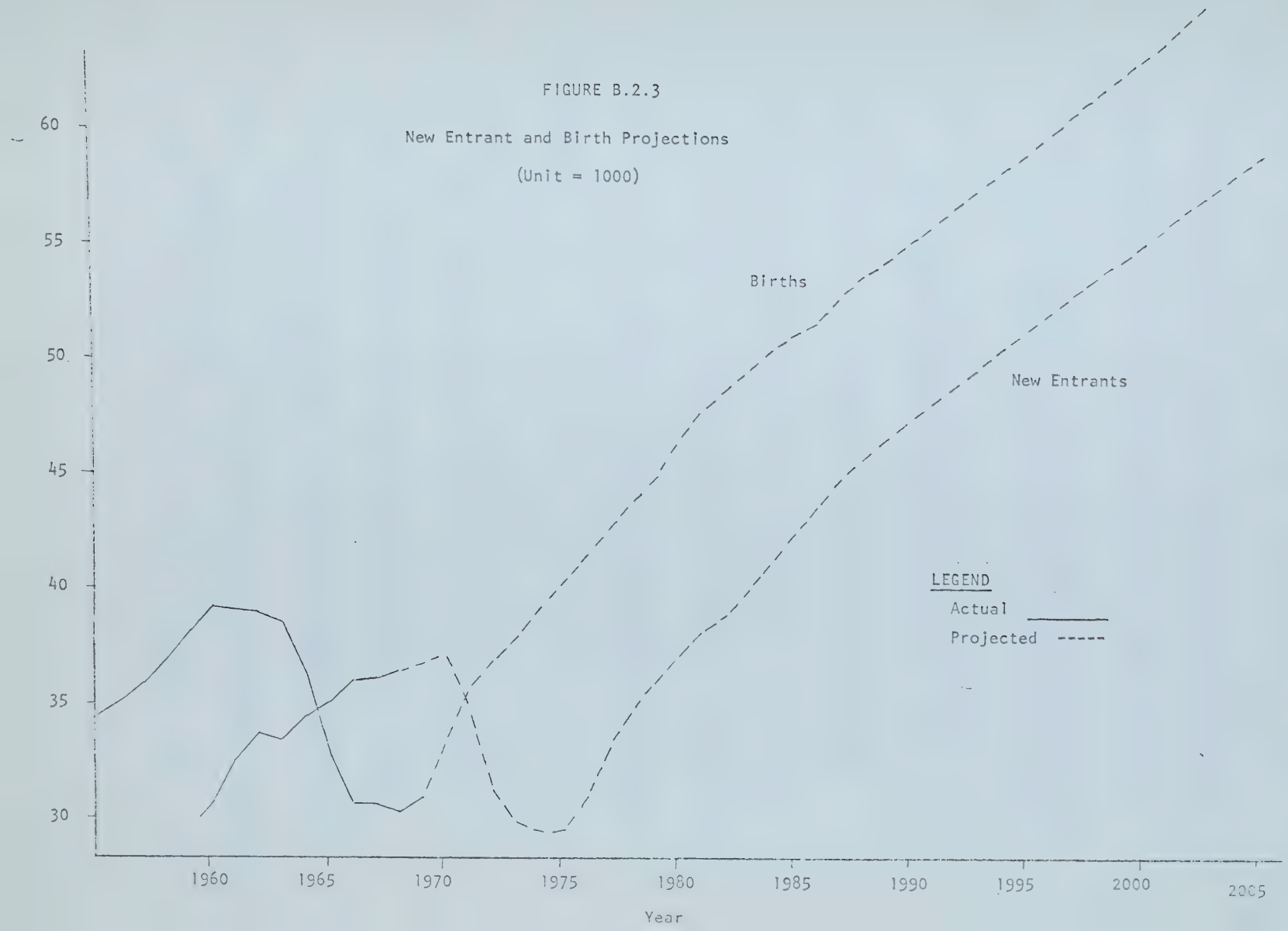
New Entrants

Year	Male	Female	Combined
1970-71	18,146	17,165	35,311
1971-72	16,080	15,190	31,270
1972-73	15,164	14,617	29,781
1973-74	15,026	14,235	29,261
1974-75	15,058	14,234	29,292
1975-76	15,618	15,078	30,696
1976-77	16,742	16,204	32,946
1977-78	17,597	17,048	34,645
1978-79	18,158	17,596	35,754
1979-80	18,710	18,131	36,841
1980-81	19,261	18,661	37,923
1981-82	19,813	19,196	39,010
1982-83	20.405	19,769	40,174
1983-84	21,029	20,369	41,398
1984-85	21,654	20,970	42,625
1985-86	22,280	21,576	43,856
1986-87	22,905	22,188	45,094
1987-88	23,405	22,676	46,082
1988-89	23,819	23,078	46,897
1989-90	24,230	23,476	47,706
1990-91	24,639	23,874	48,513
1991-92	25,050	24,272	49,322
1992-93	25,451	24,662	50,113
1993-94	25,843	25,049	50,892
1994-95	26,263	25,435	51,671
1995-96	26,628	25,822	52,449
1996-97	27,020	26,206	53,226
1997-98	27,410	26,593	54,003
1998-99	27,798	26,981	54,780
1999-00	28,187	27,370	55,556
2000-01	28,575	27,758	56,333
2001-02	28,963	28,147	57,110
2002-03	29,352	28,535	57,887
2003-04	29,740	28,923	58,664
2004-05	30,129	29,312	59,440

FIGURE B.2.3

New Entrant and Birth Projections

(Unit = 1000)



BIBLIOGRAPHY

ALBERTA Bureau of Statistics, Population, 1961 and 1966.. Forecast, 1971 - 1976 - 1981 - 1986.

ANNUAL Report, Department of Education, Province of Alberta, 1946-1970.

ARMITAGE, Peter and Cyril Smith, "The Development of Computable Models of the British Educational System and Their Possible Uses," Mathematical Models in Educational Planning, OECD, 1967.

BAISUCK, A., Construction and Evaluation of a Simulation Model for Projecting Higher Education Enrollment. Rensselaer Res. Corp., Troy, N.Y., 1969.

BARTHOLOMEW, D.J., Stochastic Models for Social Processes. New York: Wiley, 1967.

BERZTISS, "Least Square Fitting of Polynomials to Irregularly Spaced Data," Siam Review, 6, 1964.

BOX, G.E.P. and G.M. Jenkins, Time Series Analysis, Forecasting and Control. Holden-Day, 1970.

BROWN, B.W., Jr., and I.R. Savage, Methodological Studies in Educational Attendance Prediction. Minneapolis: University of Minnesota, Department of Statistics, 1960.

CHAMBERS, J.M., "A New Approach to Fitting Models to Data," Applied Statistician, No. 4, 1969.

CORREA, Hector, "A Survey of Mathematical Models in Educational Planning," Mathematical Models in Educational Planning. OECD, 1966, pp. 21-93.

CORREA, Hector, Quantitative Methods of Educational Planning. International Textbook Company, 1969.

CORREA, Hector and E. Reimer, A Simulation Model for Educational Planning in Puerto Rico. San Juan: Department of Public Instruction (Mimeographed), 1969.

DAVIS, Russell G., Planning Human Resources Development - Educational Models and Schemata. Rand McNally, 1966.

DOMINION Bureau of Statistics Publication, No. 91-201, 1970.

DRAPER, N.R. and H. Smith, Applied Regression Analysis. New York: Wiley, 1968.

DRESSEL, Paul, "Comments on the Use of Mathematical Models in Educational Planning," Mathematical Models in Educational Planning, OECD, 1966, pp. 275-290.

FELLER, W., An Introduction to Probability Theory and Its Applications. Edition II, Vol. I, New York: Wiley, 1950.

GANI, J., "Formulae for Projecting Enrollments and Degrees Awarded in Universities," Journal of the Royal Statistical Society, Series A, Vol. 126, Part 3, 1963, pp. 400-409.

GRAYBILL, F.A., An Introduction to Linear Statistical Models. Vol. I, McGraww-Hill, 1961.

HAGER, H. and C. Antle, "The Choice of the Degree of a Polynomial Model," J.R.S.S., Series B, 30, 1968.

HAGGSTROM, Gus W., "On Analyzing and Predicting Enrollments and Costs in Higher Education," Proceedings of the Social Statistics Section, J.A.S.A., 1969.

HAMELMAN, Paul W., Planning and Analysis for Higher Education: Promises and Pitfalls. Department of Business Administration, College of Business, Virginia Polytechnic Institute, Working Paper No. 8.

HOCKING, R.R. and R.N. Leslie, "Selection of the Best Subset in Regression Analysis," Vol. I, No. 4, Technometrics, Nov., 1967.

HUDSON, D.J., "Least Squares Fitting of a Polynomial Constrained to be Non-Negative, Non-Decreasing or Convex," 31, J.R.S.S., 1969, pp. 113-118.

JACOBY, E.G., Methods of School Enrollment Projections. UNESCO, Educational Studies and Documents, No. 32, 1959.

KEMMENY, J.G. and J.L. Snell, Finite Markov Chains. Toronto: D. VAn Nostrand, Co., Inc., 1969.

KENDALL, M.G. and A. Stuart, The Advanced Theory of Statistics. New York: Hafner, Vol. III.

LAMOTTE, L.R. and R.R. Hocking, "Computational Efficiency in the Selection of Regression Variables," Vol. 12, No. 1, Technometrics, Feb., 1970.

LI, W.L., "A Demographic Model of Student Progression," Proceedings of the Social Statistics Section, J.A.S.A., 1969, pp. 380-388.

LIE, Bangee Alfred, Estimating Future School Enrollment in Developing Countries: A Manual of Methodology, UNESCO, 1964.

LINS, L.J., Methodology of Enrollment Projections for Colleges and Universities. A.A.C.R.A.O., 1960.

LIU, B.A., Estimating Future School Enrollment in Developing Countries, New York: UNESCO, Population Studies No. 40, 1966.

MARSAGLIA, G., "Generating Discrete Random Variables in a Computer," Vol. VI, Communication of the A.C.M., Jan., 1963, pp. 30-38.

MARTIN, F.F., Computer Modelling and Simulation. New York: John Wiley and Sons, Inc., 1968.

MORRISON, Peter A., Implications of Migration Histories for Model Design. The Rand Corporation, Santa Monica, California, April, 1970, p. 4342.

NAYLOR, T.J., J.L. Balintfy, D.S. Burdick, and K. Chu, Computer Simulation Techniques. New York: John Wiley, 1968.

OKADA, Tetsuo, Birth and Death Projections used in Present Student-Teacher Population Growth Models. Technical Note No. 11, National Center for Educational Statistics, Division of Operations Analysis, 1966.

PERRELLA, Vera C. and Elizabeth Waldman, Out-of-School Youth -- Two Years Later. Special Labor Force Report No. 71, U.S. Department of Labor, Bureau of Labor Statistics, August, 1966.

PROJECTIONS to Educational Statistics to 1976-77. U.S. Government Printing Office, Washington, 1968.

QUENOUILLE, M.H., The Analysis of Multiple Time Series. No. 1, Griffin's Statistical Monograph and Courses, London: Charles Griffin, 1957.

ROBINSON, E.A., Multi-Channel Time Series Analysis with Digital Computer Programs. San Francisco: Holden-Day, 1967.

SAWIRIS, Milad Y., "The Projection of College Enrollment, Multivariate Behavioral Research," Volume 5, No. 1, January, 1970, pp. 83-100.

SCHATZOFF, M., R. Tsao and S. Fienberg, "Efficient Calculation of all Possible Regressions," Vol. 10, No. 4, Technometrics, November, 1968.

SHORT, Verl M., Forecasting Pupil Population. Educational Administration Center, Northern Illinois University, 1967.

STONE, Richard, "A Model of Educational Systems," Vol. III, No. 2, Minerva, Winter, 1965, pp. 172-186.

THE FORECASTING Institute of the Swedish Central Bureau of Statistics, "Projection Models of the Swedish Educational System," Mathematical Models in Educational Planning, OECD, 1966, pp. 95-123.

THONSTAD, Tore, Education and Manpower: Theoretical Models and Empirical Application. University of Toronto Press, 1968.

U.S. Bureau of Census. Projections of School and College Enrollment in the United States to 1985. P-25, No. 388, Current Population Reports, 1966.

U.S. Bureau of Census, Revised Projections of School and College Enrollment in the United States to 1985. P-25, No. 365, Current Population Reports, 1967.

VALIAHO, Hannu, "A Synthetic Approach to Stepwise Regression Analysis," Vol. 34, No. 12, Commentationes Physico-Mathematicae, 1969.

WATSON, Cicely, Saeed Quazi and Aribert Kleist, Ontario Secondary School Enrollment Projections to 1981-82. Projection, 1969.

WOLD, Herman, "Forecasting by the Chain Principle," New York: John Wiley & Sons, Inc., Time Series Analysis - Edited by M. Rosenblatt, 1963, pp. 471-497.

WURTELE, Zivia S., Mathematical Models for Educational Planning, System Development Corporation. SP-3015, November, 1967.

ZABROWSKI, Edward K., Methodology Used in Estimate First-Stage Elements of the Transition Probability Matrices for Dynamod II: Teachers and Extra-System Flows. Technical Note, National Center for Educational Statistics, 1967. Office of Education, U.S.A.

ZABROWSKI, E.K. and J.T. Hudman, Drop-Out and Retention Rate Methodology Used to Estimate First-Stage Elements of the Transition Probability Matrices for Dynamod II. Technical Note No. 28, National Center for Educational Statistics, Office of Education, U.S.A.

ZINTER, Judith R., Estimation of Age Transition Probabilities, National Center for Educational Statistics, Division of Operations Analysis, Technical Note No. 12, Office of Education, U.S.A., December 8, 1966.

ZINTER, Judith R., Estimation of Second-Stage Drop-Out Rates for Student-Teacher Population Growth Model (Dynamod II). National Center for Educational Statistics, Division of Operations Analysis, Office of Education, U.S.A., Technical Note No. 40, August 28, 1967.

ACCO PRESS
GENUINE PRESSBOARD BINDER
1150
ACCO CANADIAN COMPANY LTD.
TORONTO CANADA